Exploration-based learning of a stabilizing controller predicts locomotor adaptation
Reviewer #1 (Remarks to the Author):

I found this to be a very well-written manuscript that presents significant and novel findings about the nature of gait adaptation specifically, and motor learning more generally. The authors propose a hierarchical theory for learning optimal locomotor patterns and validate the model predictions with a series of experiments from previous studies. Their results convincingly demonstrate that this framework (stabilizing feedback controller as the inner loop + learning and memory formation as the outer loop) explains well adaptations observed when people change speeds, walk on treadmills with belts that move at different speeds, walk with loads placed asymmetrically on the legs, and walk with assistance provided by exoskeletons. I suspect that the neuromechanics community—which includes neuroscientists, roboticists, and clinicians—will find these results novel and important. Indeed, I think it will likely become a classic for its new insights, and rigour in testing them. While this paper has many strengths, there are areas that I believe need further attention from the authors.

Major comments:

I would like to see where this theory is falsifiable. In other words, what are the possible adaptation scenarios that cannot be explained by this hierarchical theory of learning? What new experiments would most rigorously test the authors’ assertions? This belongs best in the Discussion section. More generally, I recommend the Discussion section also include a discussion of the limitations of this advancement.

Although the hierarchical theory of learning makes sense, the manuscript would benefit from greater elaboration on this topic. What is different in this theory compared to previous learning theories? What is the existing biological evidence for hierarchical learning/control? Has this been hypothesized and tested for other motor systems? This belongs best in the Introduction.

My biggest technical confusion is about the nature of the stabilizing feedback controller. First, were the same initial gains used for all conditions? If so, can you report those gains in the manuscript? It is remarkable to me that the same controller produces stable (but not optimal) walking for large belt speed differences, for large speed changes, for large magnitude exoskeleton assistance, etc. How is it so robust? Can you also explain why it was unnecessary to change the feedback gains (i.e. you only learned new feedforward parameters)? It was not intuitive to me. (Extended Data Fig 9, last sentence, suggests that the controller is automatically changed depending on context. I’m confused!).

I think that there are aspects of the minimal biped that are unclear. Here are some of the things I wondered: Why are there no leg swing dynamics? Why is step length and not hip torque controlled? What is the equation for metabolic cost, including the cost of leg swing and the swing leg weighting term? How are the leg swing costs calculated when there are no leg swing dynamics?
Minor comments:

In the case of added mass (figure 2b), it is evident that when adding symmetry to energy optimality as the performance criteria, the model predictions match better to experimental data. Can we say that energy optimality alone cannot be generalized to all motor tasks? This result should also be mentioned in lines 89-93 or the paragraph of line 107.

Also for the speed-changing experiment (figure 2c), it seems that the predicted results when there is learning (compared to no learning or memory (i.e. just with the feedback controller)) capture the transient section of experimental data. The authors mention that “We find that the default controller alone is able to capture the direction of the fast transients, with the slow transients captured by the reinforcement learner with memory” but this needs more clarification.

The result of the exoskeleton experiment should refer to different graphs in figure 2d. At the moment it is not clear why the model predictions match the empirical data.

The savings, generalization, and anterograde non-interference phenomena are only tested and seen in split-belt treadmill walking. What about the other tasks?

What was the rationale for the range of noise magnitude? (i.e. 1e-3 to 3e-3)

What is the rationale for the reinforcement learning updating control parameters p on each stride and not each step? Also by learnable parameters, do they mean nominal u and s?

Line 341: what is the basis for the assumption that changing K is not necessary? How would the results change if the RL had to learn K?

Line 359- I think this section on “Asymptotic gradient estimate” needs several clarifications and improvements. First, for the lay audience, the authors should explain why the gradient is a linear function in general (maybe by mentioning Taylor expansion or using first-order partial derivative equations and mapping that to equation 4). Also, shouldn’t J in equation 4 be the gradient of J instead? Second, more clarification on including the initial condition in the gradient is required? Does it mean, for example, energy cost (or its change?) depends on the initial condition in that stride? If yes, why? Third, why not use a discount rate? What would be its effect? Fourth, more clarification on not using the Bellman equation and instead using the adaptive or iterative control-like framework is needed. What are the differences? Fifth, why is there a need for the linear internal model, i.e. equation 5? the dynamics of the biped (which I guess is nonlinear) will determine the initial condition at the next step/stride and its dependence on the state at this step/stride and the parameters.
Title: I think the title is not informative enough and is not comprehensible by a lay audience. I suggest using terms to clarify “greedy improvement”, and “safe control”

Line 12- the definition of many of these terms such as versatility, and robustness to uncertainty is not clear. Maybe providing some references for each or defining them. Also how are these terms related to the specific goals of this study?

Line 15- it seems the behavioural, algorithmic and neural levels are referring to Marr’s levels of analysis. It is not clear how references 1-3 relate to this.

Line 17: I don’t understand how reference 7 is related to maintaining stability as being the objective function in the human locomotor control system.

Line 20- When referring to a framework and paradigm in research, I recommend one cite several important papers in this area and not just one.

Line 33 - Even after looking up the definition of “violable”, I’m still unclear if it is the right word. Perhaps this would be clearer if a different word was chosen.

Line 34: Is reference 7 relevant? This is an observational clinical study not a theoretical or computational study on the importance of stability or noise in sensorimotor control.

Line 43- “We characterized this default controller via high-throughput locomotor data19–21 in steady unchanging conditions, to use as the lowest level of our hierarchical model.” Please clarify this content to make it more understandable.

Line 52-53- I suggest using a simpler term than “oracular”. Also what is “unlikely to be biologically plausible”? Having or not having an oracular knowledge?

Lines 71-72 - You describe three performance metrics at the end of this paragraph leaving me wondering which was used in the next paragraph. I recommend clarifying immediately that it is just energy being used here. It is not clarified until Line 91.

Lines 77-80- Please define negative and positive step length asymmetry here.

Line 107-112- I recommend that the authors be clear that they are taking issue with the idea of “sensory prediction error” driving adaptation and cite relevant studies. In my opinion, it reduces clarity to not use the term preferred by that field.

Line 184: I suggest that more clarification on the term “robust” is needed.

Figure 1- figure 1a second to the last plot: what model was used to calculate the metabolic rate and how is the VO2 measure calculated? It’s kind of buried in the supplementary methods.

Figure 1- figure 1c is not quite clear to me. What are the axes? Specifically, the figure does not appear to show the message the authors are trying to convey: “Exploratory noise in the controller parameters allows the learner to estimate the gradient of the performance metric”. Also, figure 1d caption would benefit from being more informative. What are all the parameters? I think the figures should stand alone to some extent.

I think the orders of the sections in figure 2 should be the same as the results discussed. Maybe switching parts c and d?

There is a typo in the Extended Data Fig. 5 adaptation ‘b’: “First adaptation (red) and re-adaptation (blue)”. I suspect red and blue should be swapped.

Extended Data Fig 6 - Are there experiments where interference does exist to which you can test the model?

Extended Data Fig. 7 needs legends to clarify different graphs in the plot. Also, what does the horizontal dotted line represent?

Figure 3e: the two rightmost graphs need more clarifications for the red and blue colours used.
Reviewer #2 (Remarks to the Author):

In this study, the authors used a model-based approach to explain patterns of locomotor adaptation observed in several prior studies. Their modeling framework relied on three key elements: a stabilizing control policy that was derived from human behavioral data, a reinforcement algorithm that updates controller parameters to minimize a cost function, and a memory store that uses contextual information to update the control policy toward previously used policies if the context has been experienced in the past. The authors demonstrated that they could reproduce key features of adaptation to walking on a split-belt treadmill, adaptation to asymmetric limb loads, and adaptation to assistance provided by an exoskeleton among other tasks. Although the paper is written well and the methods are technically rigorous, the degree to which this work provides new insights about human locomotion is limited. Specifically, the paper is pitched as proposing a theory of locomotor learning, but the core of the framework is consistent with long-held ideas about the role of trial and error during motor learning. If the objective is to establish a theory of how people learn, the authors should clearly demonstrate that the proposed learning algorithm is better than other plausible alternatives. To this end, it would be helpful if the authors clarified for each task what hypotheses have been put forth in the literature, what existing evidence supports or refutes those hypotheses, and what their model adds to this evidence base.

Although the authors attempt to differentiate their approach from similar work in deep reinforcement learning, they neglect the fact that these two approaches start from a completely different baseline. Deep methods are often used to derive policy from scratch, but it’s not at all clear how the humans from whom the stabilizing control policies were derived learned their baseline policies. By seeding the simulations with a data-driven, stable solution the authors bypass the very challenge (searching for policies in high-dimensional space) that leads deep methods to adopt non-biological solutions.

Specific Points
When describing each of the simulation results, it would be helpful if the authors described the corresponding cost that the learner is seeking to minimize.
There are several terms that, if defined, would improve the readability of the manuscript. (Line 33) What is an “almost-violable” constraint? (Line 51) What would constitute a non-greedy gradient-descent strategy? (Line 52) What is “oracular knowledge” and how would this knowledge be arrived at through automatic differentiation?

Lines 101 – 103: Because split-belt walking involves the adaptation of both motor commands and perception of belt speed (Vazquez et al., 2015), it’s not immediately clear what signals the nervous system would use to know that tied belt condition has been experienced in the past. Does the proposed learning model assume that the signals used to code context are independent of the sensory recalibration that occurs during adaptation?

Vazquez A, Statton MA, Busgang SA, Bastian AJ. Split-belt walking adaptation recalibrates sensorimotor estimates of leg speed but not position or force. J Neurophysiol. 2015 Dec;114(6):3255-67. doi: 10.1152/jn.00302.2015.

Lines 108 – 109: What is the rationale for including a penalty for asymmetry? Is this a heuristic choice to better match experimental data?

Lines 362 – 364: Is this necessarily true for all J?

Line 424: I suggest moving this and the subsequent section to the beginning of the methods. Then, the two sections on stabilizing control could be combined to reduce redundancy.

Extended Data Figure 3: Shouldn't these landscapes be 3D surfaces since, for example, one can achieve a given step length asymmetry using a range of feasible stride lengths (sum of step lengths)?

Reviewer #3 (Remarks to the Author):

In this manuscript, the authors offer a computational model that reproduces behavioral phenomena observed in several different previously published experiments involving human gait. In short, the model combines a feedback controller for stability, a reinforcement learner to facilitate gait responses to a variety of different perturbations, and memory for storing and modifying acquired movement patterns. The manuscript is clear and well-written, and I commend the authors on the effort that must have been required to perform this study.

This work is related to recent work that has considered reinforcement learning models for explaining energy optimization in human walking (e.g., Selinger et al., J Exp Biol, 2019) and other work that has considered exploration variability/noise as a potential driver of optimization/reinforcement learning (e.g., Wu et al., Nat Neurosci, 2014; Therrien et al., Brain, 2016; Wong et al., J Neurophysiol, 2019; Selinger et al., J Exp Biol, 2019; Abram et al., Curr Biol,
The current study expands beyond this work largely by adding a memory component to explain longer-term phenomena observed in previous studies.

My central question is whether a unified hierarchical framework for locomotor learning is biologically appropriate. Decades of work outside of the locomotor domain (e.g., upper extremity movements, saccadic eye movements) has established that humans have access to many different learning mechanisms with distinct (and sometimes overlapping) neural substrates. As one example involving patients with neurologic deficits, persons with cerebellar dysfunction can learn via reinforcement learning (Therrien et al., Cerebellum, 2021) but not error-based adaptation (Therrien et al., Brain, 2016). Similar delineation between learning mechanisms has also been demonstrated behaviorally and computationally in healthy adults (e.g., Izawa and Shadmehr, PLoS Comput Biol, 2011).

While the work discussed above involves behaviors outside of the locomotor domain, there is also significant evidence that there are distinct learning mechanisms available to humans during locomotion. For example, work done by Amy Bastian’s group has shown that cerebellar damage causes impairment in split-belt treadmill adaptation (Morton and Bastian, J Neurosci, 2006) while these patients are able to change their gait in other ways (e.g., they can control and respond to quick changes in belt speed; Statton et al., Cerebellum, 2019). Further work in animal models by Megan Carey and colleagues has demonstrated that there are dissociable circuits for spatial and temporal components of split-belt treadmill adaptation (Darmohray et al., Neuron, 2019). Findings in persons with hemispherectomy also show that different aspects of locomotor adaptation are dissociable (Choi et al., Brain, 2009).

The consideration of split-belt treadmill walking, exoskeleton use, etc under a unitary umbrella of locomotor learning leads to challenges for the current model to predict behavioral phenomena not included in the current study:

1. Split-belt treadmill adaptation is essentially mandatory (assuming no neurologic deficit) once the walker steps onto the treadmill. There is no exploration or instruction required, and it is very uncommon for a healthy participant to fail to adapt to the perturbation. Furthermore, participants reliably show aftereffects following sufficient split-belt adaptation. However, it is very common for the walker to fail to adapt to a robotic exoskeleton without guided exploration (Selinger et al., J Exp Biol, 2019; Simha et al., 2021) regardless of the magnitude of the energetic gradients (Simha et al., 2021). In cases where participants do indeed change their gait in response to the exoskeleton, aftereffects are not mandatory (as in the exoskeleton considered in this manuscript, Ochoa et al., J Neurophysiol, 2017). In sum, one cannot simply ‘turn off’ the learning that occurred during split-belt treadmill walking, but it does appear that this is possible following the use of an exoskeleton. It is not clear to me how the current model framework could explain why split-belt treadmill adaptation is much more reliable than exoskeleton adaptation, or how people could learn to use an exoskeleton but show no aftereffects.

2. Previous studies of split-belt treadmill walking have shown that changes in behavior are not required for learning. In a study where the experimenters restricted changes in foot placement
during split-belt treadmill walking, the authors found that the participants showed aftereffects that were not different from a group that adapted in an unconstrained manner (Long et al., J Neurophysiol, 2016). Similarly, a previous study showed that leading the participant toward the eventual solution (i.e., providing visual feedback to promote a gait with more symmetric step lengths) did not result in accelerated learning. Rather, once the feedback was removed, the participants reverted to a more asymmetric gait that would have been expected had they never received the visual at all (Roemmich et al., Curr Biol, 2016). That study also showed similar patterns of behavior during the washout phase. Given that the current model relies on changes in behavior to reduce energy cost, it would be difficult for this model to explain these phenomena.

3. The savings results reproduced in Extended Data Figure 7 also highlight a challenge for the model. The model accurately reproduces better savings following an abrupt perturbation vs. a gradual perturbation, but it incorrectly predicts better savings following the extended gradual learning perturbation vs. the abrupt perturbation. In the cited paper (Roemmich and Bastian, J Neurophysiol, 2015), the opposite is true. The proposed model will show stronger motor memories if the ‘dose’ of learning is larger, but the cited paper shows that this is not necessarily true and suggests that multiple learning mechanisms are likely involved. Other previous models of motor adaptation offer a more parsimonious account of these savings results (Herzfeld et al., Science, 2014).

The idea that larger baseline exploration variability leads to better learning is enticing but needs to be tested directly. Previous results have relied largely on correlational or coincidental findings. It would be more powerful to modulate a participant’s baseline variability (without using explicit exploration of the potential solutions) to show that someone who was previously a non-optimizer can become an optimizer through an injection of baseline variability. More generally, the addition of prospectively collected experiments that are designed specifically to test predictions of the current model would strengthen the study.

While there is clear evidence that energy optimization plays an important role in human locomotion, I urge caution in determining that it must be the driving factor in all situations where energy cost decreases over time. Energy cost also decreases over time during reach adaptation (Huang et al., J Neurosci, 2012) despite the fact that reaching to the target is clearly the primary performance goal. Further work will be needed to reconcile the roles of energetics, stability, and other potential mechanisms at play (e.g., neural response to different types of prediction error) in human locomotion.
Response to reviewer comments and suggestions.

We thank you all for the kind consideration and for your thoughtful feedback on our manuscript: “Greedy improvement of safe control predicts locomotor learning across tasks and timescales”.

First, we are pleased that all three reviewers had positive things to say about the manuscript, that our manuscript was clear, well-written, and technically rigorous, with one of the reviewers suggesting that our study will become a classic and another reviewer commending the enormous scale of work involved in the study. We are also grateful that all three reviews were detailed and thoughtful, and have greatly helped us improve the paper, both in terms of suggesting additional calculations and helping us better explain our findings, while leaving our core results unchanged. We have addressed every single comment, suggestion, or critique in the three reviews in one or more of the following ways:
- performing substantial new calculations that compare our model with other adaptation mechanisms posited in the vast and frequently non-overlapping arm reaching and locomotor adaptation literature, demonstrating the scientific advance herein more clearly [resulting in 5 new figures and 5 additional pages in the main manuscript],
- revising writing that previously resulted in reviewer misunderstanding, and adding additional clarification sentences to prevent such misunderstanding in the future,
- motivating and discussing our work in the context of prior work and contrasting the assumptions and results thereof,
- showing how certain phenomena that were posed as challenges to our model could be explained or accommodated by the model,
- providing testable predictions for prospective experiments, and
- improving unclear sentences and providing additional technical details

In the following document, we address point-by-point all the comments of the reviewers, and directly paste the sections of the paper where changes have been made to address reviewer comments. We quote the reviewer remarks in bold black text and our remarks are in blue regular text.

Reviewer 1.

I found this to be a very well-written manuscript that presents significant and novel findings about the nature of gait adaptation specifically, and motor learning more generally. The authors propose a hierarchical theory for learning optimal locomotor patterns and validate the model predictions with a series of experiments from previous studies. Their results convincingly demonstrate that this framework (stabilizing feedback controller as the inner loop + learning and memory formation as the outer loop) explains well adaptations observed when people change speeds, walk on treadmills with belts that move at different speeds, walk with loads placed asymmetrically on the legs, and walk with assistance provided by exoskeletons. I suspect that the neuromechanics community—which includes neuroscientists, roboticists, and clinicians—will find these results novel and important. Indeed, I think it will likely become a classic for its new insights, and rigour in testing them.

Thank you for this summary.
Major comments:

I would like to see where this theory is falsifiable. What are the possible adaptation scenarios that cannot be explained by this hierarchical theory of learning? What new experiments would most rigorously test the authors’ assertions? This belongs best in the Discussion section. More generally, I recommend the Discussion section also include a discussion of the limitations of this advancement.

Thank you for these comments. We have now added the following paragraph in the discussion section that address the issue of prospective experiments to examine and potentially falsify untested assumptions, and to improve the model via further targeted experiments.

We have tested our model against a wide variety of adaptation studies, providing broad empirical support for the model’s predictive ability. Future work can involve the design of targeted experiments to test the different components of this model (e.g., performance metric, adaptation algorithm), as these components contain heretofore untested assumptions about locomotor adaptation. Here, we have compared the predictive ability of performance metrics such as energy, symmetry, and sensory prediction error, determining what each can predict when acting alone. Future experiments can systematically manipulate the energy landscape, sensory feedback (e.g., vision), and unforeseen perturbations during adaptation to delineate how these performance metrics are traded off by the human nervous system — our model, which allows these adaptation mechanisms to act simultaneously, can provide a framework for interpreting such experiments. Here, we have shown the sufficiency of exploration-based gradient estimation and gradient descent with a fixed learning rate in predicting diverse adaptation phenomena. Future experiments can compare the predictions of gradient descent versus alternative descent or adaptation algorithms (e.g., gradient descent with momentum or learning rate adaptation) in long timescale trials that either have gradually time-varying conditions or alternate between different conditions at various switching frequencies. Such prospective experiments would allow us to characterize the relation between the adaptation direction in experiment and the model-predicted gradient directions, thus helping to modify the model to capture a broader range of experiments. Future work can also test the generality of our framework to other motor adaptation tasks, including the model’s ability to explain savings, generalization, interference, non-learning, and other important phenomena; this application of our model to other motor tasks will require appropriate modifications to the dynamical model and the default controller.

In addition to this paragraph, we have now added in the Discussion section multiple other paragraphs with recommendations for future experiments and further testing of the model. Limitations are interwoven through these and other paragraphs in the Discussion section. Please see the added paragraphs below.

Predictive models of motor learning such as the one proposed here could be used to improve motor learning in the real world. We have made predictions about conditions that may degrade or accelerate learning consistent with prior experiments. Given this, future hypotheses for improving learning could be tested computationally within our modeling framework before testing via prospective experiments. Such experiments may either provide further evidence supporting the model or information that could help improve the model. If the goal is to improve learning to use a device (such an exoskeleton or a treadmill), the device parameters and their sequencing can be optimized in simulation to reduce the time duration to learning steady state.

Our focus has been on capturing qualitative phenomena and we did not obtain a quantitative fit by minimizing the error between model predictions and experiment. Consistent with this preference, we used a simple biped model with simplified actuation, sensing, and default controller structure, which was sufficient for broad qualitative predictions but may limit ability to produce detailed quantitative predictions; indeed, model simplicity may be a sound reason to not seek quantitative fits. While we have captured a wide variety of experimental phenomena from diverse labs, future work could use a higher dimensional musculoskeletal and sensorimotor model and test it against other prior experimental data not considered here in addition to the aforementioned prospective experiments. In these future studies, we would seek quantitative fits to many aspects of the experimentally observed adaptation behavior (e.g., detailed kinematics, kinetics, energetics, variability), not just the time course of one or two variables (as is typical) and without experiment-specific parameter tuning.
Although the hierarchical theory of learning makes sense, the manuscript would benefit from greater elaboration on this topic. What is the existing biological evidence for hierarchical learning/control? Has this been hypothesized and tested for other motor systems? This belongs best in the Introduction.

Thank you for the suggestion. We have added the following paragraph to address neural evidence for hierarchical control and more generally, the relation between our findings and the neural underpinnings of motor learning.
Our model of locomotor learning is hierarchical and modular. Evidence for the hypothesis of hierarchical and modular motor control goes back to hundred year old experiments in which decerebrate cats produced coordinated repetitive movements but not goal-directed movements\textsuperscript{73}. It is thought that fast timescale motor responses may be mediated in part by spinal circuits while longer-timescale control, adaptation, and context-dependent responses may be achieved by interaction of the cerebellum and motor-related areas of the cerebral\textsuperscript{3,74,75}. In our model, we have separated the fast timescale feedback controller and the slow timescale adaptation mechanisms into distinct interacting modules, so that damage to just the slow timescale adaptation module in the model could still preserve the fast timescale stabilizing response. Such preservation of fast timescale response to treadmill speed changes with degraded slow adaptation to a split-belt condition was found in subjects with cerebellar damage\textsuperscript{76,77}. Indeed, such studies have established that one locus of such slow timescale motor adaptation, especially involving sensory recalibrations and internal model change, is the cerebellum\textsuperscript{3,11,12,76–79}. Thus, while our model is meant to be at the Marr level 1 and 2 (computational and algorithmic levels)\textsuperscript{90}, it could inform interpretation of data on neural underpinnings. Conversely, neural data may allow us to fine-tune our model architecture: for instance, modules in the model may contain sub-modules responsible for distinct aspects of behavior which may be neurally dissociable (e.g., spatial and temporal slow adaptation\textsuperscript{56,79}). Some studies have suggested preservation of ‘reinforcement learning’ despite cerebellar ataxia\textsuperscript{11,13,81}, but such studies examined learning from explicit visual or auditory feedback, which is distinct from the implicit reinforcement learning we have proposed for energy optimization.

What is different in this theory compared to previous learning theories?

Thank you. We have now added multiple new sections and paragraphs in the manuscript to address other learning theories: in the Introduction, Results, Methods, and Discussion.

In the Introduction, we have added the following paragraphs that introduce and discuss other learning models.

Theories of motor adaptation have often been tested in the context of discrete episodic tasks such as reaching with the arm\textsuperscript{11–13}. Motor adaptation principles are adequate for such episodic tasks may not be sufficient for continuously cascading and stability-critical motor tasks such as locomotion, multi-fingered manipulation, and indeed many activities of daily living, including sports. In episodic reaching-like tasks, wherein the initial arm state is re-set at the end of each episode, the errors during one episode do not dynamically propagate to the next episode. In contrast, in continuously cascading tasks, deviations in movement patterns can have short-term and long-term dynamical consequences. For instance, the human body is top-heavy and has unstable inverted pendular dynamics\textsuperscript{14–17}, so the deviations on one step can grow exponentially leading to destabilizing consequences unless otherwise controlled. Adaptation responses that do not account for such unstable dynamics may result in undesirable instabilities or injury-causing falls\textsuperscript{18}. Prior accounts of locomotor adaptation\textsuperscript{2,9,19} ignored the interaction with such step-to-step dynamics and the resulting propagation of errors. For instance, metabolic energy reduction-based accounts\textsuperscript{2,20} simplified adaptation to be a univariate optimization process occurring without interaction with the locomotor dynamics, thus treating locomotion as being episodic akin to reaching. Similarly, error-based learning\textsuperscript{11–13}, when applied to locomotor adaptation\textsuperscript{5,10,19,21,22}, has been used without interaction with locomotor dynamics; these models fit a single or dual rate equation to the kinematic error, without considering how these errors interact with dynamics or affect stability. Here, we put forth a model that explains how humans can adapt continuously during walking while maintaining dynamic stability — without falling, despite external perturbations and sensory noise, or without needing to stop and re-start locomotion.
In the Results section, we have added two new sub-sections, copy-pasted below, as well as expanding a paragraph about minimizing asymmetry that was already present into a sub-section. These sub-sections compare with alternative existing hypotheses for motor adaptation (minimizing asymmetry, proprioceptive realignment, minimizing kinematic task error) with the same modeling framework.

Comparing to minimizing asymmetry.

To describe split-belt adaptation, researchers have treated the left-right asymmetry in step length as the error being corrected, fitting equations with one or two time constants to describe the decrease in this asymmetry. Here, we examined what predictions our model makes if step length asymmetry is used explicitly in our optimization framework as the only performance metric. We find that minimizing asymmetry does not capture the slow timescale transients in either tied-belt or split-belt locomotion. First, for changing treadmill speeds during tied belt locomotion, our model predicts fast timescale changes due to the default controller, but no further slow timescale changes due to the reinforcement learner with an asymmetry-minimizing objective. This is because when the gait is already symmetric, minimizing asymmetry produces no further changes. During split-belt adaptation, minimizing just step length asymmetry, our model predicts convergence to pure step length symmetry (Extended Fig. 13). This is in contrast to recent experiments which suggest eventual convergence to positive step length asymmetry. More generally, minimizing asymmetry is insufficient as the lone performance metric in an optimization framework, as perfect symmetry admits infinitely many locomotion patterns and minimizing asymmetry does not result in isolated local minima required for stereotypy. Thus, minimizing asymmetry alone cannot predict the many steady state locomotor phenomena predicted by minimizing energy during normal locomotion. As a corollary, when placed in any symmetric situation with a symmetric body (e.g., slopes or bilaterally symmetric exoskeletons), minimizing asymmetry will result in zero slow timescale adaptation of the controller even if the mechanical environment is changed substantially, in contrast to experimental findings. While minimizing asymmetry alone may not explain diverse locomotor phenomena, minimizing a weighted combination of energy and asymmetry, with a small weight on the asymmetry, retains the qualitative predictions of minimizing energy, while providing a further weighting parameter to fit quantitatively to data better (Fig. 2a-b and Extended Fig. 13). We did not seek quantitative fits by tuning parameters in this study; future experiments could delineate the extent to which humans have symmetry as an explicit objective in addition to energy, given that energy already penalizes large asymmetries.
My biggest technical confusion is about the nature of the stabilizing feedback controller. First, were the same initial gains used for all conditions? If so, can you report those gains in the manuscript? It is remarkable to me that the same controller produces stable (but not optimal) walking for large belt speed differences, for large speed changes, for large magnitude exoskeleton...
assistance, etc. How is it so robust? Can you also explain why it was unnecessary to change the feedback gains (i.e. you only learned new feedforward parameters)? It was not intuitive to me. (Extended Data Fig 9, last sentence, suggests that the controller is automatically changed depending on context. I’m confused!).

Thank you for this question. Indeed, we use the same default controller with the same set of parameters as the initial starting point for all conditions (and the reinforcement learner is free to modify the parameters). We have now made this clear early in the manuscript, when we introduce the notion of the initial default controller as quoted below.

We now provide the gains of the feedback controller in the Supplementary Information as follows, and described in greater detail in the Supplementary Text:

| Feedback gain parameter $K_{ij}$ | Value | Feedback gain parameter $K_{ij}$ | Value |
|----------------------------------|-------|----------------------------------|-------|
| $K_{11}$                         | -0.5211 | $K_{21}$                         | 0.2793 |
| $K_{12}$                         | -0.0859 | $K_{22}$                         | -0.0290 |
| $K_{13}$                         | -0.0073 | $K_{23}$                         | 0.0020 |
| $K_{14}$                         | -1.0559 | $K_{24}$                         | -0.0750 |

Table 1: Parameters of the default feedback controller. Here, $K_{ik}$ is defined as $\Delta u_i/\Delta s_k$, where $u$ is the control and

Please note that the default controller remains the same, but the controller changes during learning due to different experiences.

Regarding why it was not necessary to change the feedback gains, we have the following sentences, now revised:

control and state values as well as the feedback gains. In this study, we only allow the nominal values $p = [u_{\text{nominal}}, s_{\text{nominal}}]$ to change during learning. This is because there is a one-to-one mapping between these nominal or feedforward terms and the overall gait pattern changes we are trying to predict. We keep the feedback gains $K$ fixed, as the primary role of the feedback term is to keep the system stable despite fast timescale perturbations away from the current gait pattern. Given the robustness of the controller to substantial perturbations,\textsuperscript{16, 17} this stabilizing role is satisfied by fixed feedback gains $K$. Indeed, as assumed, we find that changing them is not necessary for the major phenomena discussed herein; allowing just the feedforward term to change\textsuperscript{3} is sufficient (e.g., Fig. 2a-d). Allowing the feedback gains $K$ to change may be necessary for even more stability-challenging perturbations, where the robustness of the default controller no longer is sufficient — such changes to $K$ can be accomplished with the same framework but would require incorporating the locomotor task constraints explicitly into the performance metric (e.g., not drifting off a finite treadmill, not falling, traveling a certain distance), as otherwise the feedback gains may be chosen in a manner that makes the walker unstable. During learning, we allow the $u_{\text{nominal}}$ and $s_{\text{nominal}}$ to change independently for the left and right steps, enabling adaptation to asymmetric conditions.

Finally, the confusing remark in caption to Extended Fig. 9 did not belong there and was also incorrectly phrased, so we have deleted this sentence. “The initial transient is due to the change in the constant or feedforward term in the controller, which is different for different split conditions.” It was meant to be in a different figure, previously Extended Fig. 8, now coincidentally Extended Fig. 9 again.
So we have now rephrased this sentence follows and moved it to the new Extended Fig 9 caption, as highlighted below:

I think that there are aspects of the minimal biped that are unclear. Here are some of the things I wondered: Why are there no leg swing dynamics? Why is step length and not hip torque controlled? What is the equation for metabolic cost, including the cost of leg swing and the swing leg weighting term? How are the leg swing costs calculated when there are no leg swing dynamics?

Thank you for this question. The appropriateness of these simplifying modeling choices were established in prior articles including some of our own. We have now added the following additional sentences to provide the necessary rationale in the Supplementary Information. As mentioned above, we have also added some sentences in the Discussion regarding how model simplicity may create opportunities for future work.
Leg swing dynamics. In the aforementioned model description, the swing legs had no explicit dynamics coupled to the rest of the body, but were directly controlled by the feedback controller by specifying the next foot placement. This is a simple modeling choice with a rich tradition of explaining a variety of locomotor phenomena [11, 17, 8]. Despite the lack of explicit leg-swing dynamics, the model has a leg swing cost that is computed from the distance and duration between foot placements, as mentioned in the previous paragraph on metabolic cost.

Equations for leg swing costs are as follows:

For the leg swing cost, we initially considered two versions of the model: (1) a work-based swing cost, based on the mechanical work needed to move a point foot by a given distance in a given amount of time, starting and ending at the respective belt speeds; this cost used the same weighting of positive and negative work as for the stance cost [8, 17, 25, 16]; (2) an empirical swing cost due to Doke and Kuo [6] where the cost scales with the typical force rates required for the point-mass to traverse a given distance in a given time [17, 16, 6]. We confirmed that the qualitative step length asymmetry adaptation behavior for the split-belt adaptation was similar for the two swing costs, and then for the rest of the manuscript, used the latter force-rate-related cost, which is given by the following equation:

$$E_{\text{swing}} = c_{\text{swing}} \frac{\Delta v}{\Delta t}, \quad (5)$$

where $\Delta t$ is with the swing duration and $\Delta v$ is the change in the foot speed from stance (when the foot moves with the belt) to swing (when the foot moves with speed needed to cover the distance between the foot placements within the swing duration). The multiplier for the swing cost $c_{\text{swing}} = 0.9$ was chosen to approximate the speed-step-length relationship in normal walking (Supplementary Figure 2).

Finally, as suggested, we have provided the equations for how the total cost of a step as a function of the component costs with descriptions of each component, as quoted below.

Metabolic energy cost. The metabolic energy cost $E_{\text{step}}$ for each step is a sum of two terms, a stance leg cost $E_{\text{stance}}$ and a leg swing cost $E_{\text{swing}}$. That is,

$$E_{\text{step}} = E_{\text{stance}} + E_{\text{swing}}, \quad (3)$$

The stance leg cost captures the metabolic cost of the mechanical effort during stance, and we set it equal to

$$E_{\text{stance}} = b_{\text{pos}} W_{\text{pos}} + b_{\text{neg}} |W_{\text{neg}}|, \quad (4)$$

where $W_{\text{pos}}$ and $W_{\text{neg}}$ are, respectively, the push-off positive work and the heel-strike negative work, and $b_{\text{pos}} = 4$ and $b_{\text{neg}} = 0.83$ are reciprocals of the positive and negative work efficiencies [15, 26, 16]. For the leg swing cost, we initially considered two versions of the model: (1) a work-based swing cost, based on the mechanical work needed to move a point foot by a given distance in a given amount of time, starting and ending at the respective belt speeds; this cost used the same weighting of positive and negative work as for the stance cost [8, 17, 25, 16]; (2) an empirical swing cost due to Doke and Kuo [6] where the cost scales with the typical force rates required for the point-mass to traverse a given distance in a given time [17, 16, 6]. We confirmed that the qualitative step length asymmetry adaptation behavior for the split-belt adaptation was similar for the two swing costs, and then for the rest of the manuscript, used the latter force-rate-related cost, which is given by the following equation:

$$E_{\text{swing}} = c_{\text{swing}} \frac{\Delta v}{\Delta t}, \quad (5)$$
Minor comments:

In the case of added mass (figure 2b), it is evident that when adding symmetry to energy optimality as the performance criteria, the model predictions match better to experimental data. Can we say that energy optimality alone cannot be generalized to all motor tasks? This result should also be mentioned in lines 89-93 or the paragraph of line 107.

Thank you for the question. We have now added the following new paragraph about Minimizing Asymmetry, at the end which we discuss the nuances of the role of symmetry in improving the results (expanding on the previous narrative).

Comparing to minimizing asymmetry.

To describe split-belt adaptation, researchers have treated the left-right asymmetry in step length as the error being corrected, fitting equations with one or two time constants to describe the decrease in this asymmetry\textsuperscript{1,9,10}. Here, we examined what predictions our model makes if step length asymmetry is used explicitly in our optimization framework as the only performance metric. We find that minimizing asymmetry does not capture the slow timescale transients in either tied-belt or split-belt locomotion. First, for changing treadmill speeds during tied belt locomotion\textsuperscript{27}, our model predicts fast timescale changes due to the default controller, but no further slow timescale changes due to the reinforcement learner with an asymmetry-minimizing objective. This is because when the gait is already symmetric, minimizing asymmetry produces no further changes. During split-belt adaptation, minimizing just step length asymmetry, our model predicts convergence to pure step length symmetry (Extended Fig. 13). This is in contrast to recent experiments which suggest eventual convergence to positive step length asymmetry\textsuperscript{51}. More generally, minimizing asymmetry is insufficient as the only performance metric in an optimization framework, as perfect symmetry admits infinitely many locomotion patterns\textsuperscript{61} and minimizing asymmetry does not result in isolated local minima required for stereotypy. Thus, minimizing asymmetry alone cannot predict the many steady state locomotor phenomena predicted by minimizing energy during normal locomotion\textsuperscript{27,40,62}. As a corollary, when placed in any symmetric situation with a symmetric body (e.g., slopes or bilaterally symmetric exoskeletons), minimizing asymmetry will result in zero slow timescale adaptation of the controller even if the mechanical environment is changed substantially, in contrast to experimental findings\textsuperscript{2,8,20}. While minimizing asymmetry alone may not explain diverse locomotor phenomena, minimizing a weighted combination of energy and asymmetry, with a small weight on the asymmetry, retains the qualitative predictions of minimizing energy, while providing a further weighting parameter to fit quantitatively to data better (Fig. 2a-b and Extended Fig. 13). We did not seek quantitative fits by tuning parameters in this study; future experiments could delineate the extent to which humans have symmetry as an explicit objective in addition to energy\textsuperscript{63}, given that energy already penalizes large asymmetries\textsuperscript{61}.

We have also added the following new figure, Extended Fig 13, in addition to the comparison of energy and energy+symmetry results in Fig 2a and Fig 2b (previously only Fig 2b had this comparison).

Extended Data Fig. 13. Minimizing asymmetry. Minimizing step length asymmetry as the only objective during split-belt adaptation results in a perfectly symmetric gait in the model, which conflicts with the positive step length asymmetry found in experiment as well as when predicted by an energy minimization\textsuperscript{51}.
Also for the speed-changing experiment (figure 2c), it seems that the predicted results when there is learning (compared to no learning or memory (i.e. just with the feedback controller) capture the transient section of experimental data. The authors mention that “We find that the default controller alone is able to capture the direction of the fast transients, with the slow transients captured by the reinforcement learner with memory” but this needs more clarification.

Thank you. This is correct and we have rewritten the entire paragraph to highlight the results in more detail as suggested.

We have also modified the figure to provide an inset as follows:

**Figure 2.** Hierarchical model predicts locomotor learning in multiple task settings. a. Split-belt walking\(^6,51\), that is, with the two belts going at different speeds. Model qualitatively predicts experimental transients in step length asymmetry and metabolic energy during adaptation and de-adaptation. b. Walking with an additional mass on one foot\(^5\). Model qualitatively predicts experimental transients in step length asymmetry during adaptation and de-adaptation. c. Walking on a treadmill with abrupt speed changes every 90 seconds\(^27\). Model qualitatively predicts experimentally observed step frequency changes. Transients have a fast and slow timescale, with the fast timescale change sometimes undershooting and sometimes overshooting the steady state (red and green detail). Without learning but just the controller (gray), the fast transient is preserved but the slow transient is replaced by a (noisy) constant. d-i. Walking with an exoskeleton that provides periodic propulsive impulses\(^30\). d-ii. Stride period converges to the perturbation period, implying entrainment. d-iii. The different trajectories of impulse phase indicate starting from different initial conditions and converging to the same phase, suggesting entrainment of impulse with the gait cycle. Predictions of other experimental phenomena are in Fig. 3e and the Extended Data Figures S2-S15. All quantities are non-dimensional.
The result of the exoskeleton experiment should refer to different graphs in figure 2d. At the moment it is not clear why the model predictions match the empirical data.

Thank you for the suggestion. We have now added labels for panels within Fig 2d as (d-i), (d-ii), etc. and have referred to them in the narrative as follows. We have also edited the narrative for increased clarity.

The model predicts empirical findings of how humans adapt to exoskeleton assistance. In one experiment, humans were provided with time-periodic ankle torque impulses via a robotic exoskeleton (Fig. 2d). If the time period of these external impulses (Fig. 2d-i) was close to the human stride period, the humans changed their stride frequency to entrain to this external impulse frequency, as predicted by the model (Fig. 2d-ii). Both model and experiment show entrainment that aligns the external impulse with the transition from one step to the next (Fig. 2d-iii). This entrainment is predicted by just the default feedback

The savings, generalization, and anterograde non-interference phenomena are only tested and seen in split-belt treadmill walking. What about the other tasks?

Thank you. We have added the following sentence in the Discussion (as we are not familiar with published experiments in other locomotor tasks explicitly investigating phenomena such as savings, generalization, or anterograde non-interference):

thus helping to modify the model to capture a broader range of experiments. Future work can also test the generality of our framework to other motor adaptation tasks, including the model’s ability to explain savings, generalization, interference, non-learning, and other important phenomena; this application of our model to other motor tasks will require appropriate modifications to the dynamical model and the default controller.

What was the rationale for the range of noise magnitude? (i.e. 1e-3 to 3e-3)

Thanks for this question. We have now provided this rationale in the following sentences in the Supplementary Information, where we describe all the parameter choices.

comparison to the sensory noise (see Figure 3d in the main manuscript). The sensory noise magnitude range in Figure 3d of the main manuscript (namely, $1 \times 10^{-3}$ to $3 \times 10^{-3}$) was chosen to span a range starting from smaller than the exploratory noise to larger than the exploratory noise to illustrate how too much sensory noise may overwhelm the exploratory variability; for these results, only the relative magnitudes of the two noise sources matter (as long as both are small enough to avoid falling).

What is the rationale for the reinforcement learning updating control parameters $p$ on each stride and not each step?

Thank you. We have addressed this here, revised to answer this question more clearly.

obtained via local exploration in the neighborhood of the controller to make gradual progress toward the optimum. In this formulation, the learnable parameters $p$ are changed every stride, so that the effect of the left and the right step control on the performance metric can be experienced separately before being incorporated into the parameter change. This assumption of once-a-stride parameter change is not essential; the learning framework can be used with continuous phase-dependent control with more frequent or continuous updates of control parameters.
Also by learnable parameters, do they mean nominal \( u \) and \( s \)?

Thank you. Yes, we have now clarified this ambiguity by editing “learnable parameters” to “learnable parameters \( p \)”.

And the reviewer is correct that \( 'p' \) contains the ‘nominal u and s’.

Line 341: what is the basis for the assumption that changing \( K \) is not necessary? How would the results change if the RL had to learn \( K \)?

Thank you for this question. We have now added a few new sentences addressing this choice as follows.

the overall gait pattern changes we are trying to predict. We keep the feedback gains \( K \) fixed, as the primary role of the feedback term is to keep the system stable despite fast timescale perturbations away from the current gait pattern. Given the robustness of the controller to substantial perturbations\(^{16,17} \), this stabilizing role is satisfied by fixed feedback gains \( K \). Indeed, as assumed, we find that changing them is not necessary for the major phenomena discussed herein; allowing just the feedforward term to change\(^3 \) is sufficient (e.g., Fig. 2a-d). Allowing the feedback gains \( K \) to change may be necessary for even more stability-challenging perturbations, where the robustness of the default controller no longer is sufficient —– such changes to \( K \) can be accomplished with the same framework but would require incorporating the locomotor task constraints explicitly into the performance metric (e.g., not drifting off a finite treadmill, not falling, traveling a certain distance), as otherwise the feedback gains may be chosen in a manner that makes the walker unstable. During learning, we allow the \( u_{\text{nominal}} \) and \( s_{\text{nominal}} \)

Line 359- I think this section on “Asymptotic gradient estimate” needs several clarifications and improvements.

Thank you for these suggestions. We address them all below. To clearly show how we have edited, we break up the reviewer’s comments into their component parts below, and address them one suggestion/comment at a time.

First, for the lay audience, the authors should explain why the gradient is a linear function in general (maybe by mentioning Taylor expansion or using first-order partial derivative equations and mapping that to equation 4).

We have added better explanation of why finding a linear model helps find the gradient.
Estimating the gradient of the performance metric \( J \) with respect to the parameters \( p \) is equivalent to building a local linear model relating changes in parameters \( p \) to changes in performance \( J \). This can be understood by noting that a local linear model is the same as a first order Taylor series, and the gradient \( \nabla_x f \) of a function \( f(x) \) about \( x_0 \) appears as the coefficient of the variable \( x \) in this first order Taylor series as follows:

\[
f(x) = f(x_0) + \nabla_x f \cdot (x - x_0) + \text{higher order terms} \approx \text{some constant} + \nabla_x f \cdot x.
\]

The performance \( J \) on a given stride will not only depend on the controller parameters \( p \), but the entire system trajectory, which is uniquely determined by the initial system state and the subsequent control actions given by \( p \). So, we posit a linear model that includes dependence on both \( s_i \) and \( p_i \). On stride \( i \), if the initial state is \( s_i \), the parameters are \( p_i \), and the performance over that stride is \( J_i \), a linear model relating these quantities is given by:

\[
J_i = F s_i + G p_i + H. \tag{4}
\]

Also, shouldn’t \( J \) in equation 4 be the gradient of \( J \) instead?

This equation is as intended. The left hand side is indeed \( J \), not gradient of \( J \). The coefficient of \( p_{-i} \) (namely \( G \)) is the gradient of \( J \) with respect to \( p_{-i} \) and the coefficient of \( s_i \) (namely \( F \)) is the gradient of \( J \) with respect to \( s \). We have now made this clearer as follows.

\[
J_i = F s_i + G p_i + H. \tag{4}
\]

Here, coefficient matrix \( G \) is the gradient of the performance \( J_i \) on the current stride with respect to the learning parameters \( p_i \) and the coefficient matrix \( F \) is the gradient with respect to initial state \( s_i \). Building a linear model of \( J_i \) with respect to only \( p_i \),...

Second, more clarification on including the initial condition in the gradient is required? Does it mean, for example, energy cost (or its change?) depends on the initial condition in that stride? If yes, why?

Thank you for asking this question. Yes. This choice is very critical to the performance of the optimization for continuous tasks. We have now added the following additional sentences that should provide the rationale.

The performance \( J \) on a given stride will not only depend on the controller parameters \( p \), but the entire system trajectory, which is uniquely determined by the initial system state and the subsequent control actions given by \( p \). So, we posit a linear model that includes dependence on both \( s_i \) and \( p_i \). On stride \( i \), if the initial state is \( s_i \), the parameters are \( p_i \), and the performance over that stride is \( J_i \), a linear model relating these quantities is given by:

\[
J_i = F s_i + G p_i + H. \tag{4}
\]

Here, coefficient matrix \( G \) is the gradient of the performance \( J_i \) on the current stride with respect to the learning parameters \( p_i \) and the coefficient matrix \( F \) is the gradient with respect to initial state \( s_i \). Building a linear model of \( J_i \) with respect to only \( p_i \), ignoring the dependence on state \( s_i \) leads to incorrect gradient estimates which can result in unstable learning.

Third, why not use a discount rate? What would be its effect?

We have added the following remarks to address this question.
mean energy cost over the next few strides. We found that minimizing expected performance over just one or two strides into the future resulted in unstable learning for energy optimization. In conventional reinforcement learning\(^\text{93}\), a discount factor \(0 < \gamma < 1\) is used to modify the function minimized to \(\sum_{i=1}^{\infty} \gamma^{i-1} J_i\), which prioritizes near term performance and downweights performance in the future. We did not use such a discount factor here, but using \(\gamma \approx 1\) is analogous to the asymptotic limit we have chosen, and \(\gamma\) much less than one will give results similar to optimizing over just the next few strides.

Fourth, more clarification on not using the Bellman equation and instead using the adaptive or iterative control-like framework is needed. What are the differences?

We have deleted the remark about not using the Bellman equation, as this mathematical remark seemed digressive and not relevant to the narrative here.

Fifth, why is there a need for the linear internal model, i.e. equation 5? the dynamics of the biped (which I guess is nonlinear) will determine the initial condition at the next step/stride and its dependence on the state at this step/stride and the parameters.

Thank you for the question. We have now edited as follows:

Old version:
“To compute the long-term consequences of the parameters \(p_i\), we posit that the nervous system also maintains an internal model of the dynamics, …”

New version:
To estimate the gradient with respect to long term performance, the nervous system needs to be able to predict the future. Thus, to predict the long-term consequences of the parameters \(p_i\), we posit that the nervous system maintains an internal forward model of the dynamics, that is, how the initial state \(s_i\) and the parameters \(p_i\) for a stride affects the state at the end

We also added the word ‘future’ into the following sentence:
“Given such an internal model of the dynamics, the nervous system can estimate the future consequences of parameter changes to the steady state (by effectively simulating the internal model to steady state) …”

We have now added the following remark about why we posit linear instead of nonlinear models elsewhere:
rapid adaptation to sudden changes. Also, we use a linear internal model though the full biped dynamics are nonlinear; a linear internal model is sufficient when adaptation is gradual and the model is constantly updated to be a good approximation about the current operating point.

Title: I think the title is not informative enough and is not comprehensible by a lay audience. I suggest using terms to clarify “greedy improvement”, and “safe control”

Thank you for this suggestion. We agree and have now selected the new title to be:
Exploration-based learning of a stabilizing controller predicts locomotor adaptation

Line 12- the definition of many of these terms such as versatility, and robustness to uncertainty is not clear. Maybe providing some references for each or defining them. Also how are these terms related to the specific goals of this study?

We agree and have now deleted the following sentence, where we had used terms such as versatility, etc.

“Natural locomotion combines versatility, stability, robustness to uncertainty, and energy economy in a manner not yet matched by engineered systems.”

The manuscript now just begins with the following sentence:

Humans are able to adapt their locomotion to diverse environmental conditions and bodily changes (Fig. 1a), but the mechanisms by which such locomotor adaptation happens are not fully understood. As a consequence, while numerous

Line 15- it seems the behavioural, algorithmic and neural levels are referring to Marr's levels of analysis. It is not clear how references 1-3 relate to this.

References [1-3] were just example studies involving locomotor adaptation. So we have now re-located the references as follows:

We have removed the reference to “behavioural, algorithmic and neural levels” from the Introduction but briefly mention Marr's levels of analysis in the Discussion as follows:

sensory recalibrations and internal model change, is the cerebellum, Thus, while our model is meant to be at the Marr level 1 and 2 (computational and algorithmic levels), it could inform interpretation of data on neural underpinnings.

Line 17: I don't understand how reference 7 is related to maintaining stability as being the objective function in the human locomotor control system.

Thank you. Reference [7] was about falls in the elderly. We were using this reference to indicate the importance maintaining stability by pointing to a study that characterized falls, but have now replaced that citation with a different one [18] which is more explicitly about fall related injury.

We have also revised the introduction as follows, which cites [18]:

movement patterns can have short-term and long-term dynamical consequences. For instance, the human body is top-heavy and has unstable inverted pendular dynamics, so the deviations on one step can grow exponentially leading to destabilizing consequences unless otherwise controlled. Adaptation responses that do not account for such unstable dynamics may result in undesirable instabilities or injury-causing falls. Prior accounts of locomotor adaptation ignored the interaction with such
Line 20 - When referring to a framework and paradigm in research, I recommend one cite several important papers in this area and not just one.

Thank you. This citation was an example of ‘deep reinforcement learning’. We have now added two more related references and moved these remarks to the Discussion section instead:

starting from a known default stabilizing controller, learned under normal conditions. Our approach may lend itself to comparison with the recently popularized framework of deep reinforcement learning, which use more expressive controller approximations (deep neural networks) with orders of magnitude more parameters. These methods do not assume

Line 33 - Even after looking up the definition of “violable”, I’m still unclear if it is the right word. Perhaps this would be clearer if a different word was chosen.

This was a typo. We meant to say ‘inviolable’ (i.e., cannot be violated). We have now edited with a simpler word as follows.

Old version:
“An almost inviolable constraint on human locomotion is being stable or the avoidance of falling down, …”

New version:
“A critical constraint on human locomotion is being stable or the avoidance of falling down, despite internally generated sensorimotor noise and external uncertainties or perturbations. This requirement of not falling down in locomotion is in

Line 34: Is reference 7 relevant? This is an observational clinical study not a theoretical or computational study on the importance of stability or noise in sensorimotor control.

Thank you for the suggestion. We agree and we have removed reference 7 here and no longer cite it.

Line 43- “We characterized this default controller via high-throughput locomotor data in steady unchanging conditions, to use as the lowest level of our hierarchical model.” Please clarify this content to make it more understandable.

Thank you for the suggestion. For clarity, we have broken this sentence into two sentences and revised as follows:

structure of this default controller constrains how humans adapt to a novel situation. We characterized this default controller by modeling how humans respond to small deviations from normal walking in steady unchanging conditions using high-throughput locomotor data. This default controller is the lowest level of our hierarchical model. For all the locomotor adaptation
Line 52-53 - I suggest using a simpler term than “oracular”. Also what is “unlikely to be biologically plausible”? Having or not having an oracular knowledge?

Thank you for the suggestion. That sentence was indeed a bit cryptic and we have edited it to the following new version:

without catastrophic falls, as changing the controller too quickly can lead to instabilities. Estimating the gradient through local exploration is in contrast to other ways of obtaining the gradient, for instance, via automatic differentiation\textsuperscript{45,46}, which requires the reinforcement learner to have a perfect forward simulation of the novel circumstance, which is unlikely to be biologically plausible\textsuperscript{47}.

Compared to the old version:
as it only makes gradual changes to the control policy. This reinforcement learner does not have oracular knowledge of the gradient, as is possible via automatic differentiation\textsuperscript{12,47}, which is unlikely to be biologically plausible\textsuperscript{39}. Instead, the learner in

Lines 71-72 - You describe three performance metrics at the end of this paragraph leaving me wondering which was used in the next paragraph. I recommend clarifying immediately that it is just energy being used here. It is not clarified until Line 91.

Thank you. We have added the following note immediately after defining the performance metrics:

conditions (Fig. 2). For the reinforcement learner, we tested minimizing four performance metrics: only energy expenditure, only asymmetry (specifically, step length asymmetry, defined below), a weighted sum of energy and asymmetry, and a kinematic task error. For the results discussed below initially, we use energy expenditure alone or energy expenditure with a small step length asymmetry penalty as the performance metric; as these give qualitatively similar results, we use the latter when the performance metric is not explicitly mentioned. The minimization of other metrics, including just asymmetry, are discussed in their own separate sections later.

Lines 77-80 - Please define negative and positive step length asymmetry here.

Thank you. We have now added a clause to implicitly define what negative (and thus positive) step length asymmetry means and refer to Extended Fig 1e, which provides definitions of the different step lengths.

hours\textsuperscript{1,51,52}. Within a few strides of being exposed to this condition, humans start walking with high step length asymmetry, specifically with negative step length asymmetry\textsuperscript{4,50} – that is, the step length onto the slow belt is longer than the step length onto the fast belt (see Fig. 2a and Extended Fig. 1e). This is the fastest timescale of adaptation, sometimes called “early

Line 107-112 - I recommend that the authors be clear that they are taking issue with the idea of “sensory prediction error” driving adaptation and cite relevant studies. In my opinion, it reduces clarity to not use the term preferred by that field.

Thank you for the suggestion. As mentioned above, we now have three separate sections in the Results section on error-based learning:

- one on minimizing asymmetry
- one on minimizing task error
one on proprioceptive realignment

In these sections, we use the term ‘sensory prediction error’ and note how that concept is related to what we are minimizing.

We have also introduced the term ‘sensory prediction error’ much earlier the manuscript now, adding sentences in the Introduction as follows about ‘error-based learning’:

Minimization of different types of error\textsuperscript{11,12,23,24} (e.g., sensory prediction error, task error, proprioceptive conflict) or

---

**Line 184:** I suggest that more clarification on the term “robust” is needed.

Thank you for this suggestion. We have edited the old version:

“… robust feedback controller that responds to perturbations on a faster timescale …”

to the new version:

“… feedback controller that responds to perturbations on a faster timescale and keeps the walker stable despite the perturbations …”

---

**Figure 1:** Figure 1a second to the last plot: what model was used to calculate the metabolic rate and how is the VO\textsubscript{2} measure calculated? It’s kind of buried in the supplementary methods.

Thank you for the suggestion. We assume the reviewer means Fig 2a. In order to explain better how the VO\textsubscript{2} measure is calculated, we have added the following sentence in the Methods section:

For the metabolic cost model used, the following sentence in the Methods section addresses it:

As the reviewer suggested, we have provided the relevant mathematical methods in the Supplementary Appendix, now adding additional equations to provide a more precise description.

The metabolic rate over a stride (two steps) is the metabolic cost divided by the stride time. From this stride-wise metabolic rate $\dot{E}_{\text{stride}}$, we predict what would be measured via indirect calorimetry $\dot{E}_{VO2}$, obtained by filtering the stride-wise metabolic rate with a first order linear process with a time-constant of 42 seconds (as in [19, 12]). That is,

$$\dot{E}_{VO2} = \lambda \cdot (\dot{E}_{\text{stride}} - \dot{E}_{VO2}) \quad (6)$$

where $\lambda$ is the reciprocal of the time constant. Such predicted metabolic equivalents of indirect calorimetry is denoted as measured by VO\textsubscript{2} in Fig. 2 in the main manuscript.
Figure 1- figure 1c is not quite clear to me. What are the axes? Specifically, the figure does not appear to show the message the authors are trying to convey: “Exploratory noise in the controller parameters allows the learner to estimate the gradient of the performance metric”.

Thank you for the suggestion. We have now labeled the axes as p1 and p2, as shown below, and then call them out in the caption. The figure has now become Fig 1d.

We have revised the previously brief caption (d) into the following, explaining the panels in detail:

sensorimotor transformation with a state estimator followed by the controller. d. Reinforcement learning by mining exploratory noise to estimate gradient and improve the controller. Initially, the controller parameters $p_1$ and $p_2$ are near the optimum of the initial performance landscape (blue). When conditions change, the performance contours change (blue to orange) as does the optimum. Exploratory noise in the controller parameters, allows the learner to estimate the gradient of the performance metric and follow the negative of this gradient to improve performance. e. Memory takes in task parameters and returns the stored

Also, figure 1d caption would benefit from being more informative. What are all the parameters? I think the figures should stand alone to some extent.

Thank you for the suggestion. We agree. Fig 1d is now Fig 1e, and we have now added the following expanded caption for the new Fig 1e.

e. Memory takes in task parameters and returns the stored controller parameters $p_{\text{memory}}$ and the associated performance value $J_{\text{memory}}$. We describe how memory is used in concert with gradient-based learning. The control parameters $p_t$ are updated toward memory $p_{\text{memory}}$ when doing so improves performance (memory use); memory is updated toward the current parameters otherwise. Updates toward memory is degraded if these updates are not aligned with the gradient, and this degradation is mediated by a modified cosine tuning.
I think the orders of the sections in figure 2 should be the same as the results discussed. Maybe switching parts c and d?

Thank you for catching this. We agree. Analogous to your suggestion, we have instead switched the paragraphs in the main manuscript that describe the results in Fig 2c and Fig 2d, so that the descriptions appear in the same order as figure panels.

There is a typo in the Extended Data Fig. 5 adaptation ‘b’: “First adaptation (red) and re-adaptation (blue)”. I suspect red and blue should be swapped.

Thank you for catching this typo! We have now switched the words (red and blue) in the caption so that they reflect the corresponding curves on the figure.

Extended Data Fig 6 - Are there experiments where interference does exist to which you can test the model?

Thank you. As mentioned above, we have added the following sentence in the Discussion as we are not familiar with published experiments in other locomotor tasks explicitly investigating phenomena such as interference:

thus helping to modify the model to capture a broader range of experiments. Future work can also test the generality of our framework to other motor adaptation tasks, including the model’s ability to explain savings, generalization, interference, non-learning, and other important phenomena; this application of our model to other motor tasks will require appropriate modifications to the dynamical model and the default controller.

Extended Data Fig. 7 needs legends to clarify different graphs in the plot. Also, what does the horizontal dotted line represent?

Thank you. We have also added legends to keep tracks of the different graphs more easily, as shown below. We have deleted the horizontal dotted line (which represented zero asymmetry).
Figure 3e: the two rightmost graphs need more clarifications for the red and blue colours used.

Thank you for catching this. We have addressed this suggestion two ways:

We have also added the following legend.

- Baseline (exo off and no metronome)
- Exo on
- Metronome on and exo on

We have added the caption as follows:

energy reduction. e. Model captures experimental phenomena a, b wherein a human does not adapt to an exoskeleton that provides step-frequency-dependent assistance upon first encounter, but adapts toward the energy optimal frequency when provided with broad experience across a range of frequencies via a metronome-tracking condition. On the right two panels, blue indicates baseline condition without any assistance, red indicates exoskeleton assistance condition, and green indicates metronome-tracking condition in addition to exoskeleton assistance. In the rightmost panels, the ‘exo on’ condition (red) shows no adaptation before broad experience (green), but shows adaptation after the broad experience. All quantities are non-dimensional.

Line 346 - How is gi known? (Also, note that the line numbers have gotten messed up.)

Thank you for the suggestion. Estimation of the gradient is described in subsequent paragraphs and we now refer to where its estimation is described via the highlighted parenthetical remark below:

\[ p_{i+1} = p_i - \alpha_g (g_i), \]  

(2)

where \( g_i \) is the current gradient estimate on the \( i \)th stride (see equation 6 for how it is estimated) and \( \alpha_g \) is a scalar learning rate.
I recommend more clarification here. Why the RL algorithm here is of policy iteration?

To avoid confusion, we have now edited this sentence about policy iteration to the following statement about policy gradient method:

later separately\textsuperscript{34,92}. The proposed reinforcement learning procedure directly updates the parameters of the control policy via gradient descent, so it may be considered a variant of policy gradient reinforcement learning, where the gradient is estimated as below entirely from exploratory steps\textsuperscript{93}. Because the gradient is updated from limited and noisy data (see below), it is

Line 353- typo: “method” written twice.

Thank you for catching this typo, now fixed.

Line 356 -Please clarify this sentence.

We have edited: “In this formulation, the learnable parameters are changed every stride, so that the effect of the left and the right step control can be easily accounted for.”

to

gained via local exploration in the neighborhood of the controller to make gradual progress toward the optimum. In this formulation, the learnable parameters $p$ are changed every stride, so that the effect of the left and the right step control on the performance metric can be experienced separately before being incorporated into the parameter change. This assumption

Line 360-362- these two sentences can be combined as they convey the same meaning.

Thank you for catching this. We have now edited the second of the two sentences to better set up the sentence following it.

New version:

and the coefficient matrix $F$ is the gradient with respect to initial state $s_i$. Building a linear model of $J_i$ with respect to only $p_i$, ignoring the dependence on state $s_i$ leads to incorrect gradient estimates which can result in unstable learning. Performing gradient descent using the matrix $G$ as the gradient is equivalent to reducing the performance of a single stride $J_i$, without considering the long-term implications. Minimizing just the single-stride performance $J_i$ may result in unrealistic

Old version:

Building a linear model of $J_i$ with respect to only $p_i$, ignoring the dependence on $s_i$ leads to biased gradient estimates which can result in unstable learning. The matrix $G$ is the gradient of the performance on the current stride with respect to the learning parameters $p$, but acting on this gradient alone ignores the effect of the initial state $s_i$ on the performance. Further, minimizing just the per-stride performance $J_i$ results in unrealistic optima: for instance, turning off the actuators and falling

Line 381- the sentence needs clarification.
Thank you for catching this. We have deleted this confusing sentence, as it was a digressive remark that did not add to the narrative. Deleted sentence: “Moving the control parameters along any available descent direction is also ‘greedy’, but now in the sense of not waiting to compute the fastest descent direction.”

Line 462- (Figures 2D and 3E). Should be ‘d’ and ‘e’

Thank you. Now revised ‘Figures 2D and 3E’ to ‘Figures 2d and 3e’.

We found similar such capitalization inconsistencies elsewhere, which we have now fixed.

Line 472- Code availability- on what platform and what is the URL?

We would like to clarify that we did share the code along with the submission to the journal as Supplementary Information. However, as the reviewer points out, we do not refer to a public URL to access the code in the paper, as the code will be uploaded to a public repository (such as github) upon acceptance.

We have now edited this code availability statement as follows:

**Code availability.** Code is available as Supplementary Material with this manuscript.

Supp methods Line 45 - 42 s needs citation.

Thank you for the suggestion. We have now added two citations to support these methodological remarks:
- Selinger et al 2014 and
- Miyamoto et al 1982.

Line 53 - I suggest clarifying that sum of past foot positions is the same as position.

In the supplementary text, we have now clarified that the sum of past positions is like an integral term in a PID controller: body is over the foot \( \theta = 0 \). This controller has the structure of a discrete \( PID \) control, with feedback on position (proportional term), velocity (derivative term), and sum of positions (integral term). The feedback on the body position and the sum of past body positions help with station-keeping on the treadmill, that is, not drift off the treadmill. We use the discrete sum instead of the integral because the control is discrete and once-per step, rather than continuous. The parameters \( \dot{u}_{\text{nominal}} \), \( K_v \), and \( s_{\text{nominal}} \) for the default controller
Line 60 - I suggest revising this very complicated sentence.

Thank you for the suggestion.

Old version:
The rules for memory use ensure that using memory does not affect eventual performance improvement, deferring to gradient descent when in conflict with it and only using the stored memory when it may improve performance (Fig. 1d).

New version:
parameterized by the settings in which they were learned. This stored memory is used when encountering a setting similar to one previously encountered (Fig. 1b,e), interpolating and generalizing between settings via function approximation\(^5\). Stored memory is only used when it may improve performance and does not conflict with gradient descent (Fig. 1e). Conversely, stored memory is updated when the current controller’s performance is better than that of the motor memory. See Methods for further details.

Reviewer 2.

In this study, the authors used a model-based approach to explain patterns of locomotor adaptation observed in several prior studies. Their modeling framework relied on three key elements: a stabilizing control policy that was derived from human behavioral data, a reinforcement algorithm that updates controller parameters to minimize a cost function, and a memory store that uses contextual information to update the control policy toward previously used policies if the context has been experienced in the past. The authors demonstrated that they could reproduce key features of adaptation to walking on a split-belt treadmill, adaptation to asymmetric limb loads, and adaptation to assistance provided by an exoskeleton among other tasks.

Thank you for the summary of our manuscript. We would like to highlight that this has been the first time many of these phenomena have been predicted by any model. Further, in addition to explaining observed phenomena as the reviewer summarizes,
- we make testable predictions for future experiments [see revised results and discussion section]
- we now compare with other theories/hypotheses relating to sensorimotor adaptation [see revised results, methods, and discussion sections]
- we explain why adaptation to some devices may be easier than other devices [see elaborations in the results and discussion sections].

The paper is written well and the methods are technically rigorous.

Thank you.
The degree to which this work provides new insights about human locomotion is limited.

We respectfully disagree with this statement and apologize if our writing in the original draft did not state our novel contributions and insights clearly. To correct this impression, we have added more details in the Introduction and the Discussion sections, and additional Result sections and corresponding Figures to further clarify the new insights regarding human locomotor adaptation that our work enables.

To the best of our knowledge, no other model captures adaptation on even the simplest task of tied belt speed changes (Pagliara et al 2014), let alone explaining multiple phenomena observed in protocols of split-belt adaptation, adaptation to body changes such as added mass, and adaptation to different types of exoskeletons — all the while walking continuously without falling or stopping despite the dramatic perturbations that these novel conditions entail. We have now emphasized the importance of the stabilizing controller as an inductive bias for adaptation (as our framework proposes) multiple times in this revised manuscript. We show how our approach addresses this challenge of adaptation while remaining stable (not falling), whereas previous work has entirely sidestepped this issue. Here, we show one key new addition, Extended Figure 4, which demonstrates how learning is degraded or the biped falls when the stabilizing feedback controller is turned off.
More broadly, as the reviewer notes in the summary above, we put forth a computational model of locomotor adaptation that explains a number of locomotor phenomena — to the best of our knowledge, this is the first such model. By doing this, already in our original manuscript, we contributed a specific quantifiable objective function, controller structure, and learning algorithm that lead to predictions consistent with broad experimental data, without explicitly fitting to the data — thus, our work provides new computational insight into the principles underlying locomotor adaptation. We now explicitly highlight this contribution in the Introduction section as pasted below (see next page).

As the reviewer may agree, theoretical models that predict multiple phenomena and lead to further testable predictions, have been valuable even in much simpler tasks such as reaching (e.g., Harris and
Wolpert, 1998, Todorov and Jordan, 2003, Smith and Shadmehr 2006, Herzfeld et al 2014, Heald et al 2021), and no such broadly predictive model exists in the context of locomotor adaptation — which has important real-world relevance in rehabilitation. We have argued in the current manuscript (Introduction, Results, and Discussion sections) that it has been hard to achieve such model-based insights for locomotor adaptation due to key missing conceptual elements that our model contributes (e.g. episodic versus continuous adaptation, how to adapt despite unstable dynamics and without falling). In the revised Introduction, we have added the following paragraphs highlighting the conceptual challenges that locomotor adaptation poses for previous theories, which were primarily developed in the context of reaching.

Theories of motor adaptation have often been tested in the context of discrete episodic tasks such as reaching with the arm\textsuperscript{11–13}. Motor adaptation principles adequate for such episodic tasks may not be sufficient for continuously cascading and stability-critical motor tasks such as locomotion, multi-fingered manipulation, and indeed many activities of daily living, including sports. In episodic reaching-like tasks, wherein the initial arm state is re-set at the end of each episode, the errors during one episode do not dynamically propagate to the next episode. In contrast, in continuously cascading tasks, deviations in movement patterns can have short-term and long-term dynamical consequences. For instance, the human body is top-heavy and has unstable inverted pendular dynamics\textsuperscript{14–17}, so the deviations on one step can grow exponentially leading to destabilizing consequences unless otherwise controlled. Adaptation responses that do not account for such unstable dynamics may result in undesirable instabilities or injury-causing falls\textsuperscript{18}. Prior accounts of locomotor adaptation\textsuperscript{2,9,19} ignored the interaction with such step-to-step dynamics and the resulting propagation of errors. For instance, metabolic energy reduction-based accounts\textsuperscript{2,20} simplified adaptation to be a univariate optimization process occurring without interaction with the locomotor dynamics, thus treating locomotion as being episodic akin to reaching. Similarly, error-based learning\textsuperscript{11–13}, when applied to locomotor adaptation\textsuperscript{8,10,19,21,22}, has been used without interaction with locomotor dynamics; these models fit a single or dual rate equation to the kinematic error, without considering how these errors interact with dynamics or affect stability. Here, we put forth a model that explains how humans can adapt continuously during walking while maintaining dynamic stability — without falling, despite external perturbations and sensory noise, or without needing to stop and re-start locomotion.

Minimization of different types of error\textsuperscript{11,12,23,24} (e.g., sensory prediction error, task error, proprioceptive conflict) or metabolic energy\textsuperscript{2,7,8,25} have been posited as driving motor adaptation. However, we do not know which of these objectives dominate adaptation, how humans may trade-off these objectives under different situations, and more generally, if the minimization of these objectives can predict diverse phenomena across multiple locomotor adaptation experiments. Metabolic energy minimization can explain steady state kinematics in some cases\textsuperscript{23}, but does not in other cases\textsuperscript{26}. Moreover, the algorithm by which energy is reduced while maintaining stability is not known, and understanding the learning algorithm could explain why humans only sometimes initiate metabolic optimization, and why they often fall short of the metabolic optimum\textsuperscript{7,8}. Similarly, while error-based learning models can be fit\textsuperscript{9,10,19} to asymmetry in step length changes when walking on a split-belt treadmill\textsuperscript{3,4,6,10,19}, such descriptive models cannot make predictions without fitting to adaptation phenomena. Error-based learning also cannot, for instance, explain how humans change step frequency when the belt speed of a regular (‘tied belt’) treadmill is changed\textsuperscript{27}. To be able to generalize to any experiment other than what it was fit to, a predictive model of locomotor adaptation needs to not only be tethered to locomotor dynamics and control, but also have components that allow us to precisely define energy\textsuperscript{6,8,25}, sensory prediction error\textsuperscript{1,12}, proprioceptive conflict\textsuperscript{23–29}, etc., so that the falsifiable implications of potential adaptation mechanisms relying on these quantities can be computed and tested with experiment.

In the previous version of the manuscript, we had compared energy minimization with minimizing asymmetry — showing that energy minimization captures a broader range of phenomena. In the current version, we compare to additional performance metrics such as asymmetry, task error minimization or proprioceptive realignment in their own separate sub-sections now (please see last 1.5 pages of Results section and Extended Fig. 11 and 12). We show that minimization of these other performance metrics, previously thought to be critical to locomotor adaptation, is unable to capture experimental observations without including energy as a major objective. We show screenshots of these new sections in response to the next comment below.

In the previous version of the paper, we highlighted how our model gives insight into some unexplained experimental locomotor adaptation findings. For instance, our model provides reasons
why some experiments find adaptation while others do not — we have now expanded in the Discussion section as follows.

Our model of locomotor learning is hierarchical and modular. Evidence for the hypothesis of hierarchical and modular motor control goes back to hundred year old experiments in which decerebrate cats produced coordinated repetitive movements but not goal-directed movements. It is thought that fast timescale motor responses may be mediated in part by spinal circuits while longer-timescale control, adaptation, and context-dependent responses may be achieved by interaction of the cerebellum and motor-related areas of the cerebral cortex. In our model, we have separated the fast timescale feedback controller and the slow timescale adaptation mechanisms into distinct interacting modules, so that damage to just the slow timescale adaptation module in the model could still preserve the fast timescale stabilizing response. Such preservation of fast timescale response to treadmill speed changes with degraded slow adaptation to a split-belt condition was found in subjects with cerebellar damage. Indeed, such studies have established that one locus of such slow timescale motor adaptation, especially involving sensory recalibrations and internal model change, is the cerebellum. Thus, while our model is meant to be at the Marr level 1 and 2 (computational and algorithmic levels), it could inform interpretation of data on neural underpinnings. Conversely, neural data may allow us to fine-tune our model architecture: for instance, modules in the model may contain sub-modules responsible for distinct aspects of behavior which may be neurally dissociable (e.g., spatial and temporal slow adaptation). Some studies have suggested preservation of ‘reinforcement learning’ despite cerebellar ataxia, but such studies examined learning from explicit visual or auditory feedback, which is distinct from the implicit reinforcement learning we have proposed for energy optimization.

We argued that understanding these reasons is relevant for locomotor rehabilitation with and without assistive robots. The final paragraph of the Discussion section mentions further ways in which this model could enable engineering and biomedical applications for aiding human locomotion and mobility.

Model-based predictions of locomotor learning, such as enabled here, have applications to improving human-machine interactions including robotic prostheses and exoskeletons, making such devices intrinsically more learnable or devising protocols for accelerating their learning. Comparisons of learning in healthy and impaired human populations using our modeling framework provides a means of identifying how distinct hypothesized modules of locomotor learning may be affected, potentially informing targeted rehabilitation.

The paper is pitched as proposing a theory of locomotor learning, but the core of the framework is consistent with long-held ideas about the role of trial and error during motor learning. If the objective is to establish a theory of how people learn, the authors should clearly demonstrate that the proposed learning algorithm is better than other plausible alternatives. To this end, it would be helpful if the authors clarified for each task what hypotheses have been put forth in the literature, what existing evidence supports or refutes those hypotheses, and what their model adds to this evidence base.

We thank the reviewer for encouraging us to compare our model to existing sensorimotor adaptation theories developed for reaching tasks — in the revised manuscript, we explain why previous motor adaptation theories, put forth as general theories of motor adaptation but predominantly tested in simple reaching tasks, were not easily extendable to locomotion and when we extended them to locomotion using our modeling framework, did not predict critical locomotor adaptation phenomena. We elaborate and paste relevant sections below.

Regarding the reviewer’s remark about trial and error-based theories of adaptation: Trial and error as a hypothesis for motor learning has existed before many important motor learning studies of the last couple of decades (e.g., Thoroughman and Shadmehr, 2000). Nevertheless, these later studies were still highly useful for the field, as they contributed specific hypotheses about the objective function and algorithm by which error based learning proceeds (e.g., Thoroughman and Shadmehr, 2000 found
evidence for the use of particular motor primitives). Similarly, our model puts forth specific hypotheses about the objective function, controller structure, and learning algorithm underlying a wide range of locomotor adaptation phenomena.

Most learning requires trial and error, but attempting to improve locomotion via simple trial and error without a stabilizing controller as an inductive bias can result in falling or other learning instabilities. The feedback controller in our model allows safe exploration and adaptation, and turning off the stabilizing feedback while the gait is adapted results in falls or at least substantially degrades learning (Extended Fig. 4). This shows that what control policy the learning acts on determines the effectiveness and safety of the adaptation. We also found that a number of alternative choices results in falling: prioritizing energy optimization over the near future rather than over a longer time-horizon, too high a learning rate, and updating the gradient estimate too quickly. We have posited the use of exploratory variability for reinforcement learning or optimization, as also suggested in a few studies\cite{2,11,13,44}, including experimental evidence for the role of exploration in improving error-based learning\cite{2,44}. It was not known how such exploration could be implemented to adapt while walking continuously, without ignoring the locomotor dynamics, stability, and the continuous nature of locomotion (i.e. not treating each step as an independent episode). Indeed, using simple trial and error to perform optimization, for instance, using an exploration-drive search depending on just the previous step\cite{2,13}, works for episodic reaching but will result in falling or non-learning for walking with continuous locomotor dynamics. Thus, here, we have put forth a framework for predicting how humans adapt their walking to different conditions while continuing to be stable.

We have now highlighted in the Discussion how naive trial and error without prioritizing stability will not work here and the new Extended figure 4 pasted below:
We have added the following paragraph in the Introduction to explain why previous motor adaptation theories, proposed as general theories of motor adaptation but predominantly tested in episodic reaching tasks, do not apply directly to locomotion:
Further in the Introduction, we have more clearly delineated the gaps in existing modeling frameworks that our work fills. Thank you for the suggestion.

Minimization of different types of error\textsuperscript{11,12,23,24} (e.g., sensory prediction error, task error, proprioceptive conflict) or metabolic energy\textsuperscript{2,7,8,25} have been posited as driving motor adaptation. However, we do not know which of these objectives dominate adaptation, how humans may trade-off these objectives under different situations, and more generally, if the minimization of these objectives can predict diverse phenomena across multiple locomotor adaptation experiments. Metabolic energy minimization can explain steady state kinematics in some cases\textsuperscript{23}, but does not in other cases\textsuperscript{20}. Moreover, the algorithm by which energy is reduced while maintaining stability is not known, and understanding the learning algorithm could explain why humans only sometimes initiate metabolic optimization, and why they often fall short of the metabolic optimum\textsuperscript{2,7,8}. Similarly, while error-based learning models can be fit\textsuperscript{9,10,19} to asymmetry in step length changes when walking on a split-belt treadmill\textsuperscript{3,4,6,10,19}, such descriptive models cannot make predictions without fitting to adaptation phenomena. Error-based learning also cannot, for instance, explain how humans change step frequency when the belt speed of a regular (‘tied belt’) treadmill is changed\textsuperscript{27}. To be able to generalize to any experiment other than what it was fit to, a predictive model of locomotor adaptation needs to not only be tethered to locomotor dynamics and control, but also have components that allow us to precisely define energy\textsuperscript{6,8,25}, sensory prediction error\textsuperscript{11,12}, proprioceptive conflict\textsuperscript{23,28}, etc., so that the falsifiable implications of potential adaptation mechanisms relying on these quantities can be computed and tested with experiment.

To clarify these theoretical contributions of our model better as suggested by the reviewer, we now show in the Results and Discussion sections how existing theories of motor adaptation, developed primarily for reaching tasks, do not explain locomotor adaptation. We have contrasted exploration-based learning of an energy optimal controller (i.e. our theory) with other forms of error-based learning in the literature; we show that other theories of learning like:
- minimizing task error,
- reducing sensory prediction error in the form of proprioceptive conflict, or
- minimizing asymmetry
are not sufficient to explain locomotor adaptation. Three new sub-sections in the Results section and two new figures, Extended Fig. 11 and 12) have been added. The sections from the Results are copy-pasted below. We would like to note that a version of comparison to asymmetry-minimization was present in the original manuscript and we have added the other comparisons to this version.
Comparing to minimizing asymmetry.
To describe split-belt adaptation, researchers have treated the left-right asymmetry in step length as the error being corrected, fitting equations with one or two time constants to describe the decrease in this asymmetry. Here, we examined what predictions our model makes if step length asymmetry is used explicitly in our optimization framework as the only performance metric. We find that minimizing asymmetry does not capture the slow timescale transients in either tied-belt or split-belt locomotion. First, for changing treadmill speeds during tied belt locomotion, our model predicts fast timescale changes due to the default controller, but no further slow timescale changes due to the reinforcement learner with an asymmetry-minimizing objective. This is because when the gait is already symmetric, minimizing asymmetry produces no further changes. During split-belt adaptation, minimizing just step length asymmetry, our model predicts convergence to pure step length symmetry (Extended Fig. 13). This is in contrast to recent experiments which suggest eventual convergence to positive step length asymmetry. More generally, minimizing asymmetry is insufficient as the lone performance metric in an optimization framework, as perfect symmetry admits infinitely many locomotion patterns and minimizing asymmetry does not result in isolated local minima required for stereotypy. Thus, minimizing asymmetry alone cannot predict the many steady state locomotor phenomena predicted by minimizing energy during normal locomotion. As a corollary, when placed in any symmetric situation with a symmetric body (e.g., slopes or bilaterally symmetric exoskeletons), minimizing asymmetry will result in zero slow timescale adaptation of the controller even if the mechanical environment is changed substantially, in contrast to experimental findings. While minimizing asymmetry alone may not explain diverse locomotor phenomena, minimizing a weighted combination of energy and asymmetry, with a small weight on the asymmetry, retains the qualitative predictions of minimizing energy, while providing a further weighting parameter to fit quantitatively to data better (Fig. 2a-b and Extended Fig. 13). We did not seek quantitative fits by tuning parameters in this study; future experiments could delineate the extent to which humans have symmetry as an explicit objective in addition to energy, given that energy already penalizes large asymmetries.

Comparing to minimizing generalized task error
In low-dimensional adaptation tasks such as reaching to a target, the task error to be minimized is unambiguous; for instance, in reaching tasks with visuomotor rotation, the error is defined as angular distance to the reach target. However, in higher-dimensional tasks like locomotion, analogous definitions of task error as deviation from desired body kinematics is not uniquely defined. Here, we outline some possible definitions of task error for locomotion and implement them in our modeling framework: for instance, (i) the error could be deviations from desired body states that are walking speed-dependent, further, in this case, the error could be (ii) a weighted sum of the error along different sensory dimensions, with the different dimensions weighted differently. We computed our model predictions when these errors are minimized as the only performance metric (see Methods and Extended Fig. 14a) via the exploration-driven gradient descent of Fig. 1b.

The resulting predictions were not entirely consistent with experiment. We minimized this error for making predictions about split-belt adaptation with different relative weightings of the errors along different dimensions (as in (ii) above). Different relative weightings resulted in distinct behaviors, all of which fell short of fully capturing the experimental findings: the weighting that results in eventual positive step length asymmetry, as seen in experiment, corresponded to energy increase in contrast to experiments, and on the other hand, the weighting that results in monotonic energy decrease has a steady state with substantial negative step length asymmetry, again in contrast to experiments (Extended Fig. 14b-c). A purely kinematic performance metric was similarly found to not explain adaptation to exoskeletons in prior work, where subjects achieved entrainment to exoskeleton impulses or changed their walking frequency without plateauing at the unassisted walking kinematics.
Comparison with proprioceptive realignment

Proprioceptive realignment has been proposed as a potential mechanism accounting for the adaptation seen in split-belt locomotion and for reaching tasks with visuomotor perturbations. Vasquez et al. characterized the (proprioceptively) perceived speed of the legs after a split-belt adaptation, effectively finding that humans perceived the fast leg as being systematically slower than reality or the slow leg as faster than reality or both. A causal mechanism relating this sensory recalibration to motor adaptation has not previously been proposed, and a causal mathematical model can help establish whether the sensory recalibration could result in symmetry changes consistent with experiment.

We put forth a mathematical model of proprioceptive recalibration using our framework, which enables linking body dynamics, sensory feedback (both vision and proprioception), and motor action. In our model, the two legs are expected to be on the same surface and proprioceptive deviations from this sensory prediction are perceived as an error to be corrected by recalibrating proprioception, using vision to estimate the perturbation (see Methods and Extended Fig. 12a). This is a type of sensory prediction error, as it is due to a difference between the sensory feedback and what the nervous system expects. This model results in recalibrated estimates of leg speeds such that on a split-belt treadmill, the fast leg feels slower and slow leg feels faster than reality, as in experiment (Extended Fig. 12c), with the recalibration growing in time. The model produces no recalibration when walking on a tied belt, as in experiment. We incorporated this recalibrating proprioceptive sensing as input to the feedback controller without changing other aspects of the default feedback controller to predict what proprioceptive realignment alone can predict.

Proprioceptive realignment as implemented here falls short of explaining qualitative features of split-belt locomotor adaptation. Specifically, while the initial negative step length asymmetry produced by the default controller is decreased by the proprioceptive realignment, the steady state of the adaptation still has substantial negative asymmetry (Extended Fig. 12d), falling substantially short of experimentally observed symmetry and positive step length asymmetry, which is predicted by energy optimization. Interestingly, the model shows coincidental metabolic energy decrease as a result of proprioceptive realignment (Extended Fig. 12e), but this energy decrease is not accompanied by kinematic changes observed in experiment. Thus, while proprioceptive realignment could potentially be a partial cause of split belt adaptation, it does not explain all the associated adaptation phenomena, as also suggested by recent experiments. Beyond split-belt adaptation, proprioceptive realignment cannot explain how humans respond to tied belt speed changes, as experiments did not find significant proprioceptive realignment in the tied belt condition. Finally, proprioceptive realignment via interaction with vision, as implemented here, cannot explain adaptation to purely mechanical changes to the body or the environment such as an added mass or an exoskeleton.

Further, we have added the following paragraph to the Discussion section, contrasting how our model is able to capture phenomena abstract error-based learning cannot.

Our model naturally predicts the various ‘qualitative’ features of short timescale and long timescale responses to perturbations without fitting to the adaptation phenomena being explained. This is in contrast to the single rate or dual rate or memory of errors models of adaptation, which when applied to locomotor adaptation without including biodynamic control, do require fits or specific assumptions to capture the direction of both the slow and the fast timescale transients. Here, we predict the short timescale response to sudden perturbations as simply the response of the default feedback controller to those perturbations, and this prediction obtains the correct direction or sign of the response without fits to the data it tries to predict. For instance, our model naturally predicts that the immediate transient upon a split-belt perturbation or a leg mass addition is negative step length asymmetry (Fig. 2-3), not previously captured in a biomechanically grounded model. Similarly, we have shown that a substantial part of slow timescale motor adaptation can be predicted by performance optimization, with energy consumption as the performance metric. This model obtains the correct direction of the slow adaptation without any fits to the adaptation data. In contrast, in the traditional dual rate or memory of errors adaptation models, the direction of slow adaptation is toward zeroing the error and, therefore, is dictated by how error is defined. Thus, while descriptive models may be fit to short and long timescale transients in some locomotor adaptation experiments, they do not make predictions of the transients from more primitive assumptions. In addition, we have shown that many common ways of defining error, when coupled with locomotor dynamics, may result in predictions that disagree with experiments.
Regarding memory models, we have noted the following distinctions here and elsewhere in the manuscript.

Our accounting of savings is complementary to previous work that have addressed savings or other related phenomena via memory mechanisms centering on context inference for error-based learning or for performance improvement. However, these previous works did not consider the interaction of performance improvement and stabilizing control in a complex task such as locomotion, as here, or when considering locomotion, did not consider locomotor dynamics and control. Our memory model is also different from models that adapt the 'error sensitivity' (learning rate) of adaptation via a memory of sensory errors, which can capture savings in the form of faster adaptation rates, but cannot capture systematically lowered initial transients during re-adaptation after a complete washout as seen in experiment. In this respect, the memory of errors model of savings is similar to other linear time-invariant state-space models, which cannot capture savings in rate after a complete washout in addition to not being able to capture changes in initial transient.

Although the authors attempt to differentiate their approach from similar work in deep reinforcement learning, they neglect the fact that these two approaches start from a completely different baseline. Deep methods are often used to derive policy from scratch, but it's not at all clear how the humans from whom the stabilizing control policies were derived learned their baseline policies. By seeding the simulations with a data-driven, stable solution the authors bypass the very challenge (searching for policies in high-dimensional space) that leads deep methods to adopt non-biological solutions.

Thank you for the comment. We had added a contrasting remark to deep learning mainly because some recent studies aim to port deep reinforcement learning to model human locomotor control (Kidzinski et al. 2018). We have now revised this discussion of the differences with deep learning approaches to clarify the dimensions along which the two approaches are different and how they may be relevant in different regimes of learning, as the reviewer points out. We have moved the remarks from the Introduction to Discussion.

New version:

Our model demonstrates that a local exploration-based search strategy and remarkably simple linear controller structure is sufficient to describe the continuous adaptation of locomotion by human adults to changes to their body and their environment, starting from a known default stabilizing controller, learned under normal conditions. Our approach may lend itself to comparison with the recently popularized framework of deep reinforcement learning, which use more expressive controller approximations (deep neural networks) with orders of magnitude more parameters. These methods do not assume initialization with a default controller but instead employ highly exploratory search involving thousands of discrete walking episodes, often involve falling and resetting the initial condition at the end of each episode. Thus, these learning methods operate in a different regime from our model and are not aimed at explaining gradual human locomotor adaptation.

Specific Points

When describing each of the simulation results, it would be helpful if the authors described the corresponding cost that the learner is seeking to minimize.

Thank you for the suggestion. We have now added the following statement:
We used this hierarchical model in concert with a minimal model of a walking human (see Methods and Supplementary Text) to predict how humans will adapt to walking in many different novel and familiar conditions, including on a split-belt treadmill, with an asymmetrically added leg mass, with external assistance or perturbations from an exoskeleton, and on a treadmill whose speed is changed without warning. The model predicted the principal aspects of the locomotor learning in all these conditions (Fig. 2). For the reinforcement learner, we tested minimizing four performance metrics: only energy expenditure, only asymmetry (specifically, step length asymmetry, defined below), a weighted sum of energy and asymmetry, and a kinematic task error. For the results discussed below initially, we use energy expenditure alone or energy expenditure with a small step length asymmetry penalty as the performance metric; as these give qualitatively similar results, we use the latter when the performance metric is not explicitly mentioned. The minimization of other metrics, including just asymmetry, are discussed in their own separate sections later.

There are several terms that, if defined, would improve the readability of the manuscript. (Line 33) What is an “almost-violable” constraint?

This was a typo. We meant to say ‘inviolable’ (ie., cannot be violated). We have now edited with a simpler word as follows.

Old version:
“An almost inviolable constraint on human locomotion is being stable or the avoidance of falling down, …”

New version:
A critical constraint on human locomotion is being stable or the avoidance of falling down, despite internally generated

(Line 51) What would constitute a non-greedy gradient-descent strategy?

Gradient descent is indeed necessarily greedy and that is what we intended to say with the following sentence in the original manuscript: “This learner is ‘greedy’, as it follows the negative gradient, the direction of fastest improvement of the objective, …”

Greedy is a term used to characterize algorithms that prioritize short-term gain.

For improved clarity, we have also now revised the title of the manuscript to “Exploration-based learning of a stabilizing controller predicts locomotor adaptation”.

(Line 52) What is “oracular knowledge” and how would this knowledge be arrived at through automatic differentiation?

Thank you for the question. That sentence was indeed a bit cryptic and we have edited it to the following new version:

without catastrophic falls, as changing the controller too quickly can lead to instabilities. Estimating the gradient through local exploration is in contrast to other ways of obtaining the gradient, for instance, via automatic differentiation$^{45, 46}$, which requires the reinforcement learner to have a perfect forward simulation of the novel circumstance, which is unlikely to be biologically plausible$^{47}$. 

36
Compared to the old version:
as it only makes gradual changes to the control policy. This reinforcement learner does not have oracular knowledge of the gradient, as is possible via automatic differentiation\textsuperscript{12,27}, which is unlikely to be biologically plausible\textsuperscript{28}. Instead, the learner in

---

Lines 101 – 103: Because split-belt walking involves the adaptation of both motor commands and perception of belt speed (Vazquez et al., 2015), it's not immediately clear what signals the nervous system would use to know that tied belt condition has been experienced in the past. Does the proposed learning model assume that the signals used to code context are independent of the sensory recalibration that occurs during adaptation?

Vazquez A, Statton MA, Busgang SA, Bastian AJ. Split-belt walking adaptation recalibrates sensorimotor estimates of leg speed but not position or force. J Neurophysiol. 2015 Dec;114(6):3255-67. doi: 10.1152/jn.00302.2015.

Thank you for this comment and great question. Note that it seems incorrect to characterize the Vazquez et al 2015 study as commenting directly on the perception of ‘belt speed’ — as the instructions to the subject was regarding ‘perceived leg speed,’ not necessarily ‘belt speed.’ These verbal descriptions can potentially impact performance on such perceptual tasks.

We previously had the following remarks about how the belt speeds could be inferred by the nervous system from available visual and proprioceptive information as follows.

controller parameters for each walking speed and \( F_j(v_{belt}) \) outputs the corresponding performance metric value. In this case, the nervous system could infer the belt speed \( v_{belt} \) from the sensory stream\textsuperscript{41}, by fusing proprioception (which can infer the speed of head relative to foot, \( v_{head}/foot \)) and vision (which can sense the speed of head relative to lab, \( v_{head}/lab \)), so that the belt speed is given by: \( v_{belt} = v_{foot}/lab = v_{head}/lab - v_{head}/foot \). We assume that this task parameter inference is independent of any potential perceptual recalibration\textsuperscript{23,28}, which is addressed separately in a later section on proprioceptive realignment.

We have now edited this paragraph to add the following final sentence (highlighted in blue) to address the reviewer question:

controller parameters for each walking speed and \( F_j(v_{belt}) \) outputs the corresponding performance metric value. In this case, the nervous system could infer the belt speed \( v_{belt} \) from the sensory stream\textsuperscript{62}, by fusing proprioception (which can infer the speed of head relative to foot, \( v_{head}/foot \)) and vision (which can sense the speed of head relative to lab, \( v_{head}/lab \)), so that the belt speed is given by: \( v_{belt} = v_{foot}/lab = v_{head}/lab - v_{head}/foot \). We assume that this task parameter inference is independent of any potential perceptual recalibration\textsuperscript{23,28}, which is addressed separately in a later section on proprioceptive realignment.

As referenced above, we have a separate section in which sensory recalibration is modeled, and show how sensory recalibration (proprioceptive realignment) does not fully explain adaptation during split belt locomotion. Please see below.
Lines 108 – 109: What is the rationale for including a penalty for asymmetry? Is this a heuristic choice to better match experimental data?

Thank you for the question. Other motor control and biomechanics researchers had proposed the asymmetry penalty as a performance objective or the `error' that the adaptation minimizes. We show that such a cost alone is neither necessary nor sufficient to predict locomotor adaptation phenomena, compared to energy-minimization. We have now expanded the discussion related to asymmetry in its own sub-section as follows:
Comparing to minimizing asymmetry.
To describe split-belt adaptation, researchers have treated the left-right asymmetry in step length as the error being corrected, fitting equations with one or two time constants to describe the decrease in this asymmetry.\textsuperscript{1,9,10} Here, we examined what predictions our model makes if step length asymmetry is used explicitly in our optimization framework as the only performance metric. We find that minimizing asymmetry does not capture the slow timescale transients in either tied-belt or split-belt locomotion. First, for changing treadmill speeds during tied belt locomotion, our model predicts fast timescale changes due to the default controller, but no further slow timescale changes due to the reinforcement learner with an asymmetry-minimizing objective. This is because when the gait is already symmetric, minimizing asymmetry produces no further changes. During split-belt adaptation, minimizing just step length asymmetry, our model predicts convergence to pure step length symmetry (Extended Fig. 13). This is in contrast to recent experiments which suggest eventual convergence to positive step length asymmetry.\textsuperscript{31} More generally, minimizing asymmetry is insufficient as the lone performance metric in an optimization framework, as perfect symmetry admits infinitely many locomotion patterns and minimizing asymmetry does not result in isolated local minima required for stereotypy. Thus, minimizing asymmetry alone cannot predict the many steady state locomotor phenomena predicted by minimizing energy during normal locomotion\textsuperscript{27,40,62}. As a corollary, when placed in any symmetric situation with a symmetric body (e.g., slopes or bilaterally symmetric exoskeletons), minimizing asymmetry will result in zero slow timescale adaptation of the controller even if the mechanical environment is changed substantially, in contrast to experimental findings\textsuperscript{2,8,26}. While minimizing asymmetry alone may not explain diverse locomotor phenomena, minimizing a weighted combination of energy and asymmetry, with a small weight on the asymmetry, retains the qualitative predictions of minimizing energy, while providing a further weighting parameter to fit quantitatively to data better (Fig. 2a-b and Extended Fig. 13). We did not seek quantitative fits by tuning parameters in this study; future experiments could delineate the extent to which humans have symmetry as an explicit objective in addition to energy\textsuperscript{63}, given that energy already penalizes large asymmetries\textsuperscript{61}.

---

Lines 362 – 364: Is this necessarily true for all J?

Great point. Thank you for pointing this out. We have revised as follows, acknowledging that this may not be true for all J.

New version:

\[ J_i \] without considering the long-term implications. Minimizing just the single-stride performance \[ J_i \] may result in unrealistic optima for some performance metrics: turning off the actuators and falling may be optimal when only minimizing energy over one step. So, for non-transient tasks such as steady walking, we hypothesize that the human prioritizes the long term or steady

Old version:

“Minimizing just the per-stride performance \[ J_i \] results in unrealistic optima: for instance, turning off the actuators and falling may be optimal.”

---

Line 424: I suggest moving this and the subsequent section to the beginning of the methods. Then, the two sections on stabilizing control could be combined to reduce redundancy.

While we appreciate the suggestion, we ultimately preferred the slight redundancy to (1) first briefly introduce the stabilizing controller in a simplified way to inform our discussion of learning and (2) finally explain the details of the biped dynamics and the stabilizing controller on which the learning framework is implemented. While the dynamic and controller details in (2) are critical to the results of the manuscript, these details are not relevant to describing the learning algorithm.
Extended Data Figure 3: Shouldn’t these landscapes be 3D surfaces since, for example, one can achieve a given step length asymmetry using a range of feasible stride lengths (sum of step lengths)?

The reviewer is correct that this could have been a 3D surface plot, but we intentionally made it 2d for illustrative clarity. Specifically, we wished to illustrate the curvature along two separate directions (namely step length asymmetry and step time asymmetry). 3D surface plots do not tend to convey relative curvatures in orthogonal directions well. We have now acknowledged the reviewer’s remark in the caption as follows:

away from the optimum along the step length asymmetry compared to step time asymmetry. Finally, note that a given step length or step time asymmetry can be chosen in infinitely many ways, for instance, with different total or average step lengths; so this would need to be a higher dimensional plot to get a better characterization of the energy landscape. We have shown two slices of this higher dimensional cost landscape to compare the curvatures along two directions in the units chosen; the slices are chosen through the optimum, with the other free variables chosen to be their optimal value given the appropriate step length or step time asymmetry value.

Reviewer 3.

In this manuscript, the authors offer a computational model that reproduces behavioral phenomena observed in several different previously published experiments involving human gait. In short, the model combines a feedback controller for stability, a reinforcement learner to facilitate gait responses to a variety of different perturbations, and memory for storing and modifying acquired movement patterns.

Thank you summary of the model and what it was able to do.

Indeed, as the reviewer indicates, this model was able to predict locomotor adaptation phenomena in several different previously published experiments with a single computational framework without fitting to adaptation data. This helps us identify the computational components necessary to generate these locomotor adaptation phenomena, while respecting the challenge of interfacing with the unstable and continuous body dynamics. (We now highlight how our theoretical framework addresses this particular challenge of adapting while continuously walking without falling in a new Extended figure 4).

Further, in addition to explaining diverse phenomena as the reviewer summarizes, we have now:
- showed how our model can be used to make testable predictions for future experiments with some examples [see revised discussion section]
- compared with other theories/hypotheses relating to sensorimotor adaptation including minimizing asymmetry, error, and proprioceptive conflict [see revised results, methods, and discussion sections]
- explained why adaptation to some devices may be easier than other devices [please see elaborations in revised discussion section]

We especially appreciated your remark about “multiple adaptation mechanisms,” really taking it to heart. As mentioned in the above bulleted list, we have now proposed how to implement the complementary adaptation mechanisms (proprioceptive realignment, minimizing kinematic error and minimizing asymmetry) and compared their predictions to our original model, while acknowledging the importance of studying how the different mechanisms contribute to learning and adaptation.
The manuscript is clear and well-written, and I commend the authors on the effort that must have been required to perform this study.

Thank you, we appreciate the reviewer’s critical feedback which helped improved the clarity of our writing and helped better communicate the strength of our claims and novel contributions.

We especially appreciated your remark about “multiple adaptation mechanisms,” really taking it to heart: we have now proposed how to implement the complementary adaptation mechanisms (proprioceptive realignment, minimizing kinematic error and minimizing asymmetry) and compared their predictions to our original model.

This work is related to recent work that has considered reinforcement learning models for explaining energy optimization in human walking (e.g., Selinger et al., J Exp Biol, 2019) and other work that has considered exploration variability/noise as a potential driver of optimization/reinforcement learning (e.g., Wu et al., Nat Neurosci, 2014; Therrien et al., Brain, 2016; Wong et al., J Neurophysiol, 2019; Selinger et al., J Exp Biol, 2019; Abram et al., Curr Biol, 2022). The current study expands beyond this work largely by adding a memory component to explain longer-term phenomena observed in previous studies.

The reviewer is correct that our work is related to the above work, had previously cited a couple of them, and have now provided a more detailed narrative on how our work substantially differs from these, in terms of hypotheses, the tasks, the algorithms, and the results. Please see copy-pasted paragraphs below. We have also now cited these work earlier in the manuscript.

However, we believe the characterization by the reviewer that the current study expands current knowledge “largely by adding a memory mechanism” leaves out a number of other contributions in the manuscript, and we apologize if we did not highlight these contributions sufficiently in the previous draft. We have now improved the current manuscript to better highlight these contributions.

Specifically, as the reviewer noted in their summary of the paper, the following novel conceptual contributions make it possible for our model to be broadly predictive, none of which exists in these prior work, to the best of our knowledge:

1. stabilizing default controller that enables and shapes continuous adaptation without falling
2. learning algorithm that estimates the gradient through intentional exploration while interfacing with a dynamical model and a stabilizing controller
3. memory model to store and update control policies while interfacing with a dynamical model.

We believe these features are crucial for predicting contact-rich and continuously cascading motor learning tasks such as locomotion, multi-fingered manipulation, and sports, which require balancing stability with effort and other task goals. We have added the following paragraph in the Introduction to discuss why theories developed for episodic motor tasks may not be adequate for such continuously cascading tasks:
Theories of motor adaptation have often been tested in the context of discrete episodic tasks such as reaching with the arm\textsuperscript{11–13}. Motor adaptation principles adequate for such episodic tasks may not be sufficient for continuously cascading and stability-critical motor tasks such as locomotion, multi-fingered manipulation, and indeed many activities of daily living, including sports. In episodic reaching-like tasks, wherein the initial arm state is re-set at the end of each episode, the errors during one episode do not dynamically propagate to the next episode. In contrast, in continuously cascading tasks, deviations in movement patterns can have short-term and long-term dynamical consequences. For instance, the human body is top-heavy and has unstable inverted pendular dynamics\textsuperscript{14–17}, so the deviations on one step can grow exponentially leading to destabilizing consequences unless otherwise controlled. Adaptation responses that do not account for such unstable dynamics may result in undesirable instabilities or injury-causing falls\textsuperscript{18}. Prior accounts of locomotor adaptation\textsuperscript{2,9,19} ignored the interaction with such step-to-step dynamics and the resulting propagation of errors. For instance, metabolic energy reduction-based accounts\textsuperscript{2,20} simplified adaptation to be a univariate optimization process occurring without interaction with the locomotor dynamics, thus treating locomotion as being episodic akin to reaching. Similarly, error-based learning\textsuperscript{11–13}, when applied to locomotor adaptation\textsuperscript{5,10,19,21,22}, has been used without interaction with locomotor dynamics; these models fit a single or dual rate equation to the kinematic error, without considering how these errors interact with dynamics or affect stability. Here, we put forth a model that explains how humans can adapt continuously during walking while maintaining dynamic stability — without falling, despite external perturbations and sensory noise, or without needing to stop and re-start locomotion.

In the following paragraph in the Introduction:
- we make the point that other theories of motor adaptation have not been shown to have broad predictive ability and
- we describe what is needed in a generalizable modeling framework in order to be able to compare the different theories.

Minimization of different types of error\textsuperscript{11,12,23,24} (e.g., sensory prediction error, task error, proprioceptive conflict) or metabolic energy\textsuperscript{7,8,25} have been posited as driving motor adaptation. However, we do not know which of these objectives dominate adaptation, how humans may trade-off these objectives under different situations, and more generally, if the minimization of these objectives can predict diverse phenomena across multiple locomotor adaptation experiments. Metabolic energy minimization can explain steady state kinematics in some cases\textsuperscript{26}, but does not in other cases\textsuperscript{27}. Moreover, the algorithms by which energy is reduced while maintaining stability is not known, and understanding the learning algorithm could explain why humans only sometimes initiate metabolic optimization, and why they often fall short of the metabolic optimum\textsuperscript{7,8}. Similarly, while error-based learning models can be fit\textsuperscript{9,10,19} to asymmetry in step length changes when walking on a split-belt treadmill\textsuperscript{13,4,6,10,19}, such descriptive models cannot make predictions without fitting to adaptation phenomena. Error-based learning also cannot, for instance, explain how humans change step frequency when the belt speed of a regular (‘tied belt’) treadmill is changed\textsuperscript{22}. To be able to generalize to any experiment other than what it was fit to, a predictive model of locomotor adaptation needs to not only be tethered to locomotor dynamics and control, but also have components that allow us to precisely define energy\textsuperscript{6,8,25}, sensory prediction error\textsuperscript{1,12}, proprioceptive conflict\textsuperscript{23,28}, etc., so that the falsifiable implications of potential adaptation mechanisms relying on these quantities can be computed and tested with experiment.

The following paragraph from the Discussion section directly addresses the prior work on exploration-driven reinforcement learning that the reviewer mentions above (Wu et al, 2014, Therrien et al 2016, Selinger et al, 2019, etc.), and emphasizes:
- how straightforward trial and error or non-dynamics-aware reinforcement learning is not appropriate for locomotion,
- why the learning has to happen around a stabilizing controller.
- and how these earlier work did not explicitly consider these issues
To provide more specific distinctions between our work and the studies mentioned by the reviewer: Wu et al study describes experimental evidence for using intentional exploration. This study focused on an error minimization task without modeling the coupling to the intrinsic body dynamics or consider stability issues or energetics. In Thierren et al 2016, there was an exploration based reinforcement learning using information from only one step in the past and without considering dynamics or including a stabilizing controller, which was appropriate for episodic reaching but would not generalize to walking. In Selinger et al, 2019, a simple mathematical model is proposed, treating the locomotor learning as happening without interaction with locomotor dynamics and without considering the non-episodic aspects. This model cannot show, as we do, that both the initial timescale and the long-timescale behaviors can be explained by a single model. This distinction is also made in the Introduction now:

step-to-step dynamics and the resulting propagation of errors. For instance, metabolic energy reduction-based accounts\textsuperscript{2,20} simplified adaptation to be a univariate optimization process occurring without interaction with the locomotor dynamics, thus treating locomotion as being episodic akin to reaching. Similarly, error-based learning\textsuperscript{11-13}, when applied to locomotor

In Wong et al or Abrams et al, the hypothesis of exploration-driven learning is mentioned but no new mathematical model is considered.

Additionally, our model captures both fast and slow timescale locomotor adaptation phenomena (not only longer-term phenomena as the reviewer states above) across a wide range of experiments with a single model, where, by predictive, we mean without fitting to the adaptation data). We do not know of other prior work that achieves these locomotor predictions across timescales. We apologize if our writing could have made these contribution clearer, and have tried to improve in the current version.

We have now more clearly outlined our overall contributions in the final paragraph of the Introduction now.

Our contribution in this manuscript is a model which causally links body dynamics, control policy, learning algorithm, and memory for locomotion, thus showing how different adaptation hypotheses can be examined within the same modeling framework. We model adaptation as an exploration-driven gradient-based improvement of a stabilizing feedback controller, explaining how humans can learn to walk better over multiple timescales without falling or stopping. The model predicts different adaptation phenomena in over ten different prior experimental studies\textsuperscript{2,3,3-8,29,30}, heretofore not captured by a single model. We use this model to predict critical locomotor adaptation phenomena such as fast timescale reactive response followed by slow timescale adaptation, savings, faster de-adaptation, generalization, non-learning in some humans and in some situations, and ineffectiveness of gradual learning protocols, in a manner that accounts for locomotor dynamics. Such model-based prediction of human locomotor learning phenomena could help us design better rehabilitation paradigms, build assistive robots that are more learnable and provide greater benefit\textsuperscript{31,32}, and more generally, accelerate motor learning\textsuperscript{31,33}. 
As noted in the paragraph above, we have also added new sections in the results where we explicitly compare the predictions of our model to those made by other adaptation mechanisms proposed in the literature (proprioceptive alignment, asymmetry, and task error reduction). Thank you for your suggestions about the complementary adaptation mechanisms. We have added the following new sub-sections in the Results section that first show how our framework enables the implementation of other adaptation mechanisms and then describe the predictions from minimizing step length asymmetry, kinematic task error, or proprioceptive conflict.

Comparing to minimizing asymmetry.
To describe split-belt adaptation, researchers have treated the left-right asymmetry in step length as the error being corrected, fitting equations with one or two time constants to describe the decrease in this asymmetry\cite{1,9,10}. Here, we examined what predictions our model makes if step length asymmetry is used explicitly in our optimization framework as the only performance metric. We find that minimizing asymmetry does not capture the slow timescale transients in either tied-belt or split-belt locomotion. First, for changing treadmill speeds during tied belt locomotion\cite{27}, our model predicts fast timescale changes due to the default controller, but no further slow timescale changes due to the reinforcement learner with an asymmetry-minimizing objective. This is because when the gait is already symmetric, minimizing asymmetry produces no further changes. During split-belt adaptation, minimizing just step length asymmetry, our model predicts convergence to pure step length symmetry (Extended Fig. 13). This is in contrast to recent experiments which suggest eventual convergence to positive step length asymmetry\cite{27}. More generally, minimizing asymmetry is insufficient as the lone performance metric in an optimization framework, as perfect symmetry admits infinitely many locomotion patterns\cite{61} and minimizing asymmetry does not result in isolated local minima required for stereotypy. Thus, minimizing asymmetry alone cannot predict the many steady state locomotor phenomena predicted by minimizing energy during normal locomotion\cite{27,40,62}. As a corollary, when placed in any symmetric situation with a symmetric body (e.g., slopes or bilaterally symmetric exoskeletons), minimizing asymmetry will result in zero slow timescale adaptation of the controller even if the mechanical environment is changed substantially, in contrast to experimental findings\cite{2,8,26}. While minimizing asymmetry alone may not explain diverse locomotor phenomena, minimizing a weighted combination of energy and asymmetry, with a small weight on the asymmetry, retains the qualitative predictions of minimizing energy, while providing a further weighting parameter to fit quantitatively to data better (Fig. 2a-b and Extended Fig. 13). We did not seek quantitative fits by tuning parameters in this study; future experiments could delineate the extent to which humans have symmetry as an explicit objective in addition to energy\cite{63}, given that energy already penalizes large asymmetries\cite{61}.

Comparing to minimizing generalized task error
In low-dimensional adaptation tasks such as reaching to a target, the task error to be minimized is unambiguous; for instance, in reaching tasks with visuomotor rotation, the error is defined as angular distance to the reach target\cite{5,12}. However, in higher-dimensional tasks like locomotion, analogous definitions of task error as deviation from desired body kinematics is not uniquely defined. Here, we outline some possible definitions of task error for locomotion and implement them in our modeling framework: for instance, (i) the error could be deviations from desired body states that are walking speed-dependent; further, in this case, the error could be (ii) a weighted sum of the error along different sensory dimensions, with the different dimensions weighted differently. We computed our model predictions when these errors are minimized as the only performance metric (see Methods and Extended Fig. 14a) via the exploration-driven gradient descent of Fig. 1b.

The resulting predictions were not entirely consistent with experiment. We minimized this error for making predictions about split-belt adaptation with different relative weightings of the errors along different dimensions (as in (ii) above). Different relative weightings resulted in distinct behaviors, all of which fell short of fully capturing the experimental findings: the weighting that results in eventual positive step length asymmetry, as seen in experiment, corresponded to energy increase in contrast to experiments, and on the other hand, the weighting that results in monotonic energy decrease has a steady state with substantial negative step length asymmetry, again in contrast to experiments (Extended Fig. 14b-c). A purely kinematic performance metric was similarly found to not explain adaptation to exoskeletons in prior work, where subjects achieved entrainment to exoskeleton impulses\cite{30} or changed their walking frequency\cite{39} without plateauing at the unassisted walking kinematics.
My central question is whether a unified hierarchical framework for locomotor learning is biologically appropriate. Decades of work outside of the locomotor domain (e.g., upper extremity movements, saccadic eye movements) has established that humans have access to many different learning mechanisms with distinct (and sometimes overlapping) neural substrates.

Thank you for this question and comment.

First, as quoted in response to the previous comment, we now implement test multiple learning mechanisms proposed in other work within and outside the locomotor domain (minimizing error, asymmetry, proprioceptive realignment). We show that these can have some qualitative predictions similar to observed in experiment (thus complementing energy optimization) while falling short of fully explaining the locomotor adaptation phenomena. This was a great addition to the manuscript, so thank you for your remarks [new sections in Results, Discussion, and Methods].

Second, our use of the term ‘unified’ appears to have been misunderstood as ‘single neural substrate’. We apologize for any imprecise wording that may have contributed to this misunderstanding. We have removed the word ‘unified’ from the manuscript to avoid further misunderstanding. We would like to clarify that what we put forth here is a Marr level 1 and 2 i.e. computational and algorithmic level model that predicts diverse behavioral observations, and can have different underlying neural implementations — which can only be tested with neural data. Our model does not preclude distinct and overlapping neural substrates. While detailed comparison with neural data is outside the scope of
In this study, we have added the following paragraph explicitly noting that the modular structure of our model is synergistic with current neural evidence for dissociable networks/mechanisms (citing many of the studies that the reviewer pointed out), and the ways in which neural data can help evaluate the mechanisms underlying the computations involved in our model.

Our model of locomotor learning is hierarchical and modular. Evidence for the hypothesis of hierarchical and modular motor control goes back to hundred year old experiments in which decerebrate cats produced coordinated repetitive movements but not goal-directed movements. It is thought that fast timescale motor responses may be mediated in part by spinal circuits while longer-timescale control, adaptation, and context-dependent responses may be achieved by interaction of the cerebellum and motor-related areas of the cerebral cortex. In our model, we have separated the fast timescale feedback controller and the slow timescale adaptation mechanisms into distinct interacting modules, so that damage to just the slow timescale adaptation module in the model could still preserve the fast timescale stabilizing response. Such preservation of fast timescale response to treadmill speed changes with degraded slow adaptation to a split-belt condition was found in subjects with cerebellar damage. Indeed, such studies have established that one locus of such slow timescale motor adaptation, especially involving sensory recalibrations and internal model change, is the cerebellum. Thus, while our model is meant to be at the Marr level 1 and 2 (computational and algorithmic levels), it could inform interpretation of data on neural underpinnings. Conversely, neural data may allow us to fine-tune our model architecture: for instance, modules in the model may contain sub-modules responsible for distinct aspects of behavior which may be neurally dissociable (e.g., spatial and temporal slow adaptation). Some studies have suggested preservation of ‘reinforcement learning’ despite cerebellar ataxia, but such studies examined learning from explicit visual or auditory feedback, which is distinct from the implicit reinforcement learning we have proposed for energy optimization.

Regarding alternative adaptation mechanisms, we say the following in the Discussion section. Our Introduction also explicitly mentions them now.

We have tested our model against a wide variety of adaptation studies, providing broad empirical support for the model’s predictive ability. Future work can involve the design of targeted experiments to test the different components of this model (e.g., performance metric, adaptation algorithm), as these components contain heretofore untested assumptions about locomotor adaptation. Here, we have compared the predictive ability of performance metrics such as energy, symmetry, and sensory prediction error, determining what each can predict when acting alone. Future experiments can systematically manipulate the energy landscape, sensory feedback (e.g., vision), and unforeseen perturbations during adaptation to delineate how these performance metrics are traded off by the human nervous system — our model, which allows these adaptation mechanisms to act simultaneously, can provide a framework for interpreting such experiments. Here, we have shown the sufficiency of

As one example involving patients with neurologic deficits, persons with cerebellar dysfunction can learn via reinforcement learning (Therrien et al., Cerebellum, 2021) but not error-based adaptation (Therrien et al., Brain, 2016). Similar delineation between learning mechanisms has also been demonstrated behaviorally and computationally in healthy adults (e.g., Izawa and Shadmehr, PLoS Comput Biol, 2011). While the work discussed above involves behaviors outside of the locomotor domain, there is also significant evidence that there are distinct learning mechanisms available to humans during locomotion. For example, work done by Amy Bastian’s group has shown that cerebellar damage causes impairment in split-belt treadmill adaptation (Morton and Bastian, J Neurosci, 2006) while these patients are able to change their gait in other ways (e.g., they can control and respond to quick changes in belt speed; Statton et al., Cerebellum, 2019).

Thank you for asking us to comment on these important studies. These experiments (Morton and Bastian, J Neurosci, 2006 vs Statton et al., Cerebellum, 2019) are perfectly in support of the modular nature of our model.
The highlighted part below describes how our model structure is consistent with findings of Morton and Bastian, J Neurosci, 2006 versus Statton et al., Cerebellum, 2019 (references 76-77 below).

Our model of locomotor learning is hierarchical and modular. Evidence for the hypothesis of hierarchical and modular motor control goes back to hundred year old experiments in which decerebrate cats produced coordinated repetitive movements but not goal-directed movements. It is thought that fast timescale motor responses may be mediated in part by spinal circuits while longer-timescale control, adaptation, and context-dependent responses may be achieved by interaction of the cerebellum and motor-related areas of the cerebrum. In our model, we have separated the fast timescale feedback controller and the slow timescale adaptation mechanisms into distinct interacting modules, so that damage to just the slow timescale adaptation module in the model could still preserve the fast timescale stabilizing response. Such preservation of fast timescale response to treadmill speed changes with degraded slow adaptation to a split-belt condition was found in subjects with cerebellar damage. Indeed, such studies have established that one locus of such slow timescale motor adaptation, especially involving sensory recalibrations and internal model change, is the cerebellum. Thus, while our model is meant to be at the Marr level 1 and 2 (computational and algorithmic levels), it could inform interpretation of data on neural underpinnings. Conversely, neural data may allow us to fine-tune our model architecture: for instance, modules in the model may contain sub-modules responsible for distinct aspects of behavior which may be neurally dissociable (e.g., spatial and temporal slow adaptation). Some studies have suggested preservation of ‘reinforcement learning’ despite cerebellar ataxia, but such studies examined learning from explicit visual or auditory feedback, which is distinct from the implicit reinforcement learning we have proposed for energy optimization.

Later in the same paragraph, highlighted below, we comment that the reinforcement learning in the studies referenced above (Therrien et al., Cerebellum, 2021 and Izawa and Shadmehr, PLoS Comput Biol, 2011) is distinct from what we propose.

Our model of locomotor learning is hierarchical and modular. Evidence for the hypothesis of hierarchical and modular motor control goes back to hundred year old experiments in which decerebrate cats produced coordinated repetitive movements but not goal-directed movements. It is thought that fast timescale motor responses may be mediated in part by spinal circuits while longer-timescale control, adaptation, and context-dependent responses may be achieved by interaction of the cerebellum and motor-related areas of the cerebrum. In our model, we have separated the fast timescale feedback controller and the slow timescale adaptation mechanisms into distinct interacting modules, so that damage to just the slow timescale adaptation module in the model could still preserve the fast timescale stabilizing response. Such preservation of fast timescale response to treadmill speed changes with degraded slow adaptation to a split-belt condition was found in subjects with cerebellar damage. Indeed, such studies have established that one locus of such slow timescale motor adaptation, especially involving sensory recalibrations and internal model change, is the cerebellum. Thus, while our model is meant to be at the Marr level 1 and 2 (computational and algorithmic levels), it could inform interpretation of data on neural underpinnings. Conversely, neural data may allow us to fine-tune our model architecture: for instance, modules in the model may contain sub-modules responsible for distinct aspects of behavior which may be neurally dissociable (e.g., spatial and temporal slow adaptation). Some studies have suggested preservation of ‘reinforcement learning’ despite cerebellar ataxia, but such studies examined learning from explicit visual or auditory feedback, which is distinct from the implicit reinforcement learning we have proposed for energy optimization.

We have added an additional section on modeling adaptation via additional explicit feedback in response to a later comment by the reviewer but also applies to the remarks above:

Interaction with explicit feedback.

Our framework is meant to model implicit adaptation and learning, but can accommodate explicit adaptation mechanisms acting in parallel. One potential way to speed up locomotor adaptation is to provide explicit verbal instruction to the subject about the desired behavior or provide visual feedback on the error between desired and actual behavior (Extended Fig. 11a). Indeed, providing visual feedback on step length asymmetry to subjects on a split-belt treadmill and asking them to reduce this asymmetry hastened the progress toward symmetry — compared to adaptation without this feedback. Removing this visual feedback partway through adaptation results in the increased symmetry being largely wiped out, so that the asymmetry goes back approximately to where it would have been without the explicit feedback. We were able to capture this phenomenon (Extended Fig. 11c-d) by adding a separate module for explicit control that acts in parallel to the feedback controller in memory (Extended Fig. 11a), as in some prior work. This demonstration is simply to show that the implicit learner of Fig. 1b can be readily modified to accommodate explicit mechanisms without degrading the implicit learner’s performance. This demonstration also shows that kinematic behavior changes due to explicit corrections need not, by themselves, be sufficient to modify implicit learning, as seen in experiments.
Further work in animal models by Megan Carey and colleagues has demonstrated that there are dissociable circuits for spatial and temporal components of split-belt treadmill adaptation (Darmohray et al., Neuron, 2019). Findings in persons with hemispherectomy also show that different aspects of locomotor adaptation are dissociable (Choi et al., Brain, 2009).

Thank you for bringing this up. Our model is not in conflict with this evidence. We have added the following highlighted sentences, addressing these specific manuscripts in the following sentences:

Our model of locomotor learning is hierarchical and modular. Evidence for the hypothesis of hierarchical and modular motor control goes back to hundred year old experiments in which decerebrate cats produced coordinated repetitive movements but not goal-directed movements. It is thought that fast timescale motor responses may be mediated in part by spinal circuits while longer-timescale control, adaptation, and context-dependent responses may be achieved by interaction of the cerebellum and motor-related areas of the cerebrum. In our model, we have separated the fast timescale feedback controller and the slow timescale adaptation mechanisms into distinct interacting modules, so that damage to just the slow timescale adaptation module in the model could still preserve the fast timescale stabilizing response. Such preservation of fast timescale response to treadmill speed changes with degraded slow adaptation to a split-belt condition was found in subjects with cerebellar damage. Indeed, such studies have established that one locus of such slow timescale motor adaptation, especially involving sensory recalibrations and internal model change, is the cerebellum. Thus, while our model is meant to be at the Marr level 1 and 2 (computational and algorithmic levels), it could inform interpretation of data on neural underpinnings. Conversely, neural data may allow us to fine-tune our model architecture: for instance, modules in the model may contain sub-modules responsible for distinct aspects of behavior which may be neurally dissociable (e.g., spatial and temporal slow adaptation). Some studies have suggested preservation of ‘reinforcement learning’ despite cerebellar ataxia, but...

We had previously noted that the functional modules in our model could have sub-modules as highlighted below (these sentences are in the current revision as well).

We have argued that predicting human locomotor learning phenomena may require the following functional components: a stabilizing feedback controller, an optimizing reinforcement learner, a gradient estimator, a memory mechanism, and possibly a module that reduces sensory errors. Like all mathematical models of complex phenomena (famously in string theory), there may be multiple realizability: the same architectural hypothesis can be expressed in different terms, grouping some components together, dividing components into their sub-components, or have different realizations of similar function. No matter this...

The consideration of split-belt treadmill walking, exoskeleton use, etc. under a unitary umbrella of locomotor learning leads to challenges for the current model to predict behavioral phenomena not included in the current study:

1. Split-belt treadmill adaptation is essentially mandatory (assuming no neurologic deficit) once the walker steps onto the treadmill. There is no exploration or instruction required, and it is very uncommon for a healthy participant to fail to adapt to the perturbation. Furthermore, participants reliably show aftereffects following sufficient split-belt adaptation. However, it is very common for the walker to fail to adapt to a robotic exoskeleton without guided exploration (Selinger et al., J Exp Biol, 2019; Simha et al., 2021) regardless of the magnitude of the energetic gradients (Simha et al., 2021).

Thank you for this question. We agree that it is surprising that a single model can capture these superficially distinct phenomena — nevertheless, these distinct phenomena are indeed consistent with our model predictions, and we believe that this ability of our model to capture such distinct phenomena is a strength. Specifically, we showed that by changing the properties of the device that the human is interacting with i.e. the environment, we can predict the change from a learner to a non-
learner and vice versa. We believe this ability of our model to explain both learning and non-learning will be beneficial in informing rehabilitation, and we elaborate on this in the Discussion section.

We have now added the following new paragraph in the Discussion section, directly highlighting our proposed explanations of why humans adapt readily in some novel situations (e.g., split belt) and not others (e.g., some exoskeletons), and also make the falsifiable model-guided prediction regarding how to make split-belts hard to learn and exoskeletons easier to learn.

Our model suggests explanations for why humans may adapt reliably in some novel situations (for instance, during split belt walking\textsuperscript{4,6,21} and not others (for instance, some exoskeleton studies\textsuperscript{2,8,20,68}). One might wonder what principles underlie reliable adaptation in one class of devices and unreliable adaptation in another class. Our results suggest that the two classes of devices are not fundamentally different with respect to motor learning, but that the differences may be due to dynamical properties of treadmills versus current exoskeletons and their controllers. Our model suggests ways in which we can make subjects unreliable learners on split-belt treadmills and reliable learners on exoskeletons and prostheses. Specifically, our model predicts that split-belt adaptation can be slowed down or extinguished by noisy belt speed variations (Extended Fig. 16), while exoskeleton learning can be reliable if they are noise-free and delay-free (Extended Fig. 5). We also noted that many exoskeleton studies that did not show obligatory adaptation involved exoskeleton controllers that had a one step delay between human action and the device response\textsuperscript{2,8,20}. We showed that gradient descent can be substantially degraded or entirely stopped in the presence of such delays (Extended Fig. 15), whereas there can be reliable learning in exoskeletons with no delay (Extended Fig. 16); this prediction can be tested by systematically manipulating the device delay in future experiments. In summary, we suggest that humans may exhibit better adaptation to exoskeletons if the device has low noise, has simple consistent dynamics from step to step, and does not have substantial delay between human action and device response.

We have also added the following paragraph in the Results section about how response delays in current day exoskeletons can degrade learning.

Initial non-adaptation can also be explained as being due to the delay between human action and exoskeleton response in many of these exoskeleton studies. Many exoskeleton adaptation experiments with subjects that did not spontaneously adapt\textsuperscript{2,8,26} had an exoskeleton controller that provided assistance or resistance based on the subject’s previous walking step, resulting in a delay between action and energetic consequence. We showed that such delays can substantially degrade or even stop gradient descent based learning (Extended Data Fig. 15), making adaptation not obligatory. The gradient estimate is degraded due to the inductive bias of the reinforcement learner being inconsistent with such delays.

These paragraphs reference two new figures, copy-pasted below.
1) Extended Fig 16 shows how noise can degrade learning on split-belt adaptation
2) Extended Fig 15 shows how delayed device assistance can degrade gradient based optimization.
In cases where participants do indeed change their gait in response to the exoskeleton, aftereffects are not mandatory (as in the exoskeleton considered in this manuscript, Ochoa et al., J Neurophysiol, 2017). In sum, one cannot simply ‘turn off’ the learning that occurred during split-
belt treadmill walking, but it does appear that this is possible following the use of an exoskeleton. It is not clear to me how the current model framework could explain why split-belt treadmill adaptation is much more reliable than exoskeleton adaptation, or how people could learn to use an exoskeleton but show no aftereffects.

Thank you for this comment but we found that Ochoa et al., J Neurophysiol, 2017 study found the opposite result to that suggested above by the reviewer. Ochoa et al (2017) found that different subjects adapted (i.e. entrained) to the exoskeleton to different levels, do show after-effects, and the level of the aftereffects was approximately correlated to the level of adaptation. This is shown in Figure 11 of the Ochoa et al manuscript.

Moreover, while a split-belt treadmill is a single well-defined piece of equipment, exoskeletons come in numerous designs — all of which can have different kinds of after-effects. Another recent paper also shows after-effects upon extensive exposure to an exoskeleton (Poggensee et al. 2021).

That the aftereffects are related to the magnitude of adaptation was true of all our adaptation simulations. So, our results are consistent with that of Ochoa et al 2017.

2. Previous studies of split-belt treadmill walking have shown that changes in behavior are not required for learning. In a study where the experimenter restricted changes in foot placement during split-belt treadmill walking, the authors found that the participants showed aftereffects that were not different from a group that adapted in an unconstrained manner (Long et al., J Neurophysiol, 2016). Similarly, a previous study showed that leading the participant toward the eventual solution (i.e., providing visual feedback to promote a gait with more symmetric step lengths) did not result in accelerated learning. Rather, once the feedback was removed, the participants reverted to a more asymmetric gait that would have been expected had they never received the visual at all (Roemmich et al., Curr Biol, 2016). That study also showed similar patterns of behavior during the washout phase. Given that the current model relies on changes in behavior to reduce energy cost, it would be difficult for this model to explain these phenomena.

Thank you for this comment. This is about explicit or consciously mediated control in the presence of visual feedback. We have now added the following paragraph and Extended Fig 11 to show our model can accommodate interactions with such explicit learning mechanisms, reproducing the phenomena that the reviewer refers to above as difficult for the model to explain. We agree that such explicit learning was not the focus of our original model — which we should have more explicitly mentioned.

**Interaction with explicit feedback.**

Our framework is meant to model implicit adaptation and learning, but can accommodate explicit adaptation mechanisms acting in parallel. One potential way to speed up locomotor adaptation is to provide explicit verbal instruction to the subject about the desired behavior or provide visual feedback on the error between desired and actual behavior (Figure 10a). Indeed, providing visual feedback on step length asymmetry to subjects on a split-belt treadmill and asking them to reduce this asymmetry hastened the progress toward symmetry compared to adaptation without this feedback (Figure 10b). Removing this visual feedback partway through adaptation results in the increased symmetry being largely wiped out, so that the asymmetry goes back to (approximately) where it would have been without the explicit feedback. We were able to capture this phenomenon (Figure 10c-d) by adding a separate module that adds to or subtracts from the feedback controller in memory (Figure 10a), acting in parallel to it as in some prior work. This demonstration is simply to show that the explicit learner of Fig. 1b can be simply be modified to accommodate explicit mechanisms without degradation of the implicit learner’s performance. It also shows that changes in kinematic behavior due to explicit corrections need not, by itself, be sufficient to modify implicit learning, as seen in experiments.
3. The savings results reproduced in Extended Data Figure 7 also highlight a challenge for the model. The model accurately reproduces better savings following an abrupt perturbation vs. a gradual perturbation, but it incorrectly predicts better savings following the extended gradual learning perturbation vs. the abrupt perturbation. In the cited paper (Roemmich and Bastian, J Neurophysiol, 2015), the opposite is true. The proposed model will show stronger motor memories if the ‘dose’ of learning is larger, but the cited paper shows that this is not necessarily true and suggests that multiple learning mechanisms are likely involved. Other previous models of motor adaptation offer a more parsimonious account of these savings results (Herzfeld et al., Science, 2014).

The reviewer refers to two specific experimental results in the Roemmich and Bastian (2015) study, which we had compared our model against.

**Result-1.** Better savings following an abrupt perturbation vs. a gradual perturbation.
As the reviewer notes above, our model captures this result correctly.

**Result-2:** better savings following the abrupt perturbation vs. the extended gradual perturbation

Throughout the manuscript, our focus has been on explaining major ‘qualitative’ phenomena and have avoided explicit fits to data to capture ‘quantitative’ trends. It is indeed possible to tune our model parameters to capture the result that “better savings following the abrupt perturbation compared to extended gradual learning perturbation.” Our focus was on a minimal set of parameters across many experiments — our hope is that future work will use our framework to understand the quantitative aspects of these phenomena better. We have now edited the figure to add a version of quantitative results that agree with experiments, while making the point about qualitative vs quantitative results in the caption, as follows:

---

Extended Data Fig. 8. Protocols with abrupt, gradual, and extended adaptation regimes. a) We simulated the different adaptation protocols consisting of gradual and abrupt introduction of a split-belt condition, different extents of the adaptation regime, and gradual versus abrupt ending to the adaptation regime, analogous to those considered by Roemmich and Bastian. b) During first adaptation, the protocols that begin with abrupt change in belt speeds showed qualitatively different transients compared to the protocols with gradual changes in belt speeds. The qualitative differences in experimental data were predicted by the model. Two sets of model predictions are shown in separate roles for substantially different model choices, denoted model predictions-1 and model predictions-2. c) The model predicts that the second adaptation of every other protocol will have a smaller initial transient and more savings than ‘gradual’ protocol, as is qualitatively seen in the experimental data. Further, the longer duration abrupt first adaptation resulted in a smaller initial transient during second adaptation compared to a short abrupt protocol, both in the model and data. d) We have commented mainly on qualitative results that are robust to model and protocol parameter choice and have not attempted quantitative fits. Some other results — such as the relative ordering of savings between ‘abrupt’ and ‘extended gradual’ as shown — are dependent on model parameter choice; one can change this ordering either by changing model parameters, or by changing the ‘extended’ duration of the ‘extended gradual’ protocol, which is a falsifiable prediction that can be tested by further experiment.
For the second part of this comment, where the reviewer states “the model will show stronger motor memories if the dose of learning is larger”: This appears to be based on a misunderstanding of our model’s properties. Our model does not have the property of stronger motor memories if the dose of learning is larger, as the reviewer claims. We now discuss this property in the Results section as follows:

for episodic tasks\(^9,36\). Having a longer duration adaptation phase in which the perturbation grows gradually may result in less adaptation than a shorter adaptation phase in which the perturbation began abruptly and remains constant (Extended Data Fig. 8). Thus, using a longer dose of the experience does not necessarily result in better adaptation in our model — it depends on how constant or noise-free the dose is. That a longer dose of experience is not sufficient for better adaptation was also shown by our model predicting that noise-like fluctuations in belt speed can dramatically curtail adaptation (Extended Data Fig. 16), suggesting that an explicit memory of errors is not necessary for this property\(^9\).

The reviewer states that Herzfeld et al 2014 could capture such savings results more parsimoniously. We were unclear on which specific locomotor adaptation savings results the reviewer refers to.

- Herzfeld et al 2014 do not cite Roemmich and Bastian 2015 (which came later) and thus did not explain any of the experimental results first demonstrated by Roemmich and Bastian 2015. So they do not explain Result-2 in their manuscript. We have shown our model can do this if one chooses to tune parameters.
- Herzfeld et al 2014 do refer to and explain Result-1 quoted above, a version of which appeared first in Malone et al 2011. But we have shown that our model can also explain this result.
- If one is solely trying to capture the phenomenon in Result-1, Herzfeld et al 2014 can be viewed as more parsimonious as they have fewer parameters than our model, however our model captures a much broader range of locomotor adaptation phenomena. This could be because Herzfeld et al. treat walking as having no dynamics other than the changing of one variable, namely step length asymmetry — so cannot explain most of the other phenomena we predict in our study. Indeed, we specifically test the predictive ability of step length asymmetry as an error in our framework and show that it alone cannot capture many phenomena.
- Finally, as highlighted below, we note that our model is able to capture certain saving-related phenomena that cannot be explained by Herzfeld et al 2014.

Our accounting of savings is complementary to previous work that have addressed savings or other related phenomena via memory mechanisms centering on context inference for error-based learning or for performance improvement\(^2,36,52,55\). However, these previous works did not consider the interaction of performance improvement and stabilizing control in a complex task such as locomotion\(^36,52,55\). As here, or when considering locomotion, did not consider locomotor dynamics and control\(^2\). Our memory model is also different from models that adapt the ‘error sensitivity’ (learning rate) of adaptation via a memory of sensory errors\(^9\), which can capture savings in the form of faster adaptation rates, but cannot capture systematically lowered initial transients during re-adaptation after a complete washout as seen in experiment\(^56\). In this respect, the memory of errors model\(^9\) of savings is similar to other linear time-invariant state-space models\(^57\), which cannot capture savings in rate after a complete washout in addition to not being able to capture changes in initial transient\(^9\).

The idea that larger baseline exploration variability leads to better learning is enticing but needs to be tested directly. Previous results have relied largely on correlational or coincidental findings. It would be more powerful to modulate a participant’s baseline variability (without using explicit exploration of the potential solutions) to show that someone who was previously a non-optimizer can become an optimizer through an injection of baseline variability.
As explained in response to an earlier comment, we do not believe that the non-learning on exoskeletons is entirely due to too low exploration variability, as described in the following Discussion paragraph (quoted earlier). In fact, we think an interaction between exploration and the other components of our model is crucial to establish whether learning will improve. Indeed, we show that our model is able to explain both learning and non-learning phenomena due to this feature.

Our model suggests explanations for why humans may adapt reliably in some novel situations (for instance, during split belt walking\textsuperscript{6,21}) and not others (for instance, some exoskeleton studies\textsuperscript{2,8,20,68}). One might wonder what principles underlie reliable adaptation in one class of devices and unreliable adaptation in another class. Our results suggest that the two classes of devices are not fundamentally different with respect to motor learning, but that the differences may be due to dynamical properties of treadmills versus current exoskeletons and their controllers. Our model suggests ways in which we can make subjects unreliable learners on split-belt treadmills and reliable learners on exoskeletons and prostheses. Specifically, our model predicts that split-belt adaptation can be slowed down or extinguished by noisy belt speed variations (Extended Fig. 16), while exoskeleton learning can be reliable if they are noise-free and delay-free (Extended Fig. 5). We also noted that many exoskeleton studies that did not show obligatory adaptation involved exoskeleton controllers that had a one step delay between human action and the device response\textsuperscript{2,8,20}. We showed that gradient descent can be substantially degraded or entirely stopped in the presence of such delays (Extended Fig. 15), whereas there can be reliable learning in exoskeletons with no delay (Extended Fig. 16); this prediction can be tested by systematically manipulating the device delay in future experiments.

In summary, we suspect that humans may exhibit better adaptation to exoskeletons if the device has low noise, has simple consistent dynamics from step to step, and does not have substantial delay between human action and device response.

To address this specific question by the reviewer about increasing exploratory variability, we have added the following paragraph to the Discussion section:

A corollary to the prediction that lowering device noise improves learning reliability is that increasing baseline human exploratory variability compared to unresolved sensorimotor or device noise may improve learning reliability. It is an open question whether baseline exploration as used by the nervous system in implicit learning can be manipulated by an experimenter via purely external means (that is, via sensory or mechanical perturbations or other biofeedback) — in a manner that results in more reliable learning. One study that increased variability externally did not find better learning\textsuperscript{69} and another study performed a manipulation that increased both variability and learning\textsuperscript{70}. It is unclear if this increased variability specifically corresponds to increased exploration because both studies changed the sensory or the mechanical environment, which could have increased variability by increasing unresolved sensorimotor noise. Further, according to our model, such increased variability comes with a higher energetic cost at steady state\textsuperscript{2} as well as potentially higher fall risk, so future work could use our model in concert with targeted experiments to delineate how humans trade-off these competing objectives of exploration, energy, and stability.

More generally, the addition of prospectively collected experiments that are designed specifically to test predictions of the current model would strengthen the study. While there is clear evidence that energy optimization plays an important role in human locomotion, I urge caution in determining that it must be the driving factor in all situations where energy cost decreases over time. Energy cost also decreases over time during reach adaptation (Huang et al., J Neurosci, 2012) despite the fact that reaching to the target is clearly the primary performance goal. Further work will be needed to reconcile the roles of energetics, stability, and other potential mechanisms at play (e.g., neural response to different types of prediction error) in human locomotion.

Thank you for these remarks. We agree: as noted earlier and partly on account of your remark, we implemented different adaptation mechanisms (minimizing kinematic task error, proprioceptive alignment, minimizing asymmetry, explicit feedback driven learning) in addition to energy optimization to directly compare the different predictions. We find that these other mechanisms alone could not predict all the phenomena that our model could predict, showing that they cannot be solely responsible. Indeed, we do not rule them out as acting in concert with energy optimization [4 new sub-sections in Results and two new figures, quoted below at the end of the document]
In fact, we think our modeling framework made it possible to compare the predictive ability of these different adaptation mechanisms, and therefore we also suggest in the Discussion section now how our model helps design targeted prospective experiments that can explore the interplay between energetics, stability, and other mechanisms. [quoted/highlighted below]

We have tested our model against a wide variety of adaptation studies, providing broad empirical support for the model’s predictive ability. Future work can involve the design of targeted experiments to test the different components of this model (e.g., performance metric, adaptation algorithm)\textsuperscript{83}, as these components contain heretofore untested assumptions about locomotor adaptation. Here, we have compared the predictive ability of performance metrics such as energy, symmetry, and sensory prediction error, determining what each can predict when acting alone. Future experiments can systematically manipulate the energy landscape, sensory feedback (e.g., vision), and unforeseen perturbations during adaptation to delineate how these performance metrics are traded off by the human nervous system — our model, which allows these adaptation mechanisms to act simultaneously, can provide a framework for interpreting such experiments. Here, we have shown the sufficiency of exploration-based gradient estimation and gradient descent with a fixed learning rate in predicting diverse adaptation phenomena.

The individual sections on minimizing kinematic task error, proprioceptive alignment, minimizing asymmetry are quoted below.

**Comparing to minimizing asymmetry.**

To describe split-belt adaptation, researchers have treated the left-right asymmetry in step length as the error being corrected, fitting equations with one or two time constants to describe the decrease in this asymmetry\textsuperscript{7,9,10}. Here, we examined what predictions our model makes if step length asymmetry is used explicitly in our optimization framework as the only performance metric. We find that minimizing asymmetry does not capture the slow timescale transients in either tied-belt or split-belt locomotion. First, for changing treadmill speeds during tied belt locomotion\textsuperscript{57}, our model predicts fast timescale changes due to the default controller, but no further slow timescale changes due to the reinforcement learner with an asymmetry-minimizing objective. This is because when the gait is already symmetric, minimizing asymmetry produces no further changes. During split-belt adaptation, minimizing just step length asymmetry, our model predicts convergence to pure step length symmetry (Extended Fig. 13). This is in contrast to recent experiments which suggest eventual convergence to positive step length asymmetry\textsuperscript{51}. More generally, minimizing asymmetry is insufficient as the lone performance metric in an optimization framework, as perfect symmetry admits infinitely many locomotion patterns\textsuperscript{61}, and minimizing asymmetry does not result in isolated local minima required for stereotypy. Thus, minimizing asymmetry alone cannot predict the many steady state locomotor phenomena predicted by minimizing energy during normal locomotion\textsuperscript{27,40,62}. As a corollary, when placed in any symmetric situation with a symmetric body (e.g., slopes or bilaterally symmetric exoskeletons), minimizing asymmetry will result in zero slow timescale adaptation of the controller even if the mechanical environment is changed substantially, in contrast to experimental findings\textsuperscript{2,8,26}. While minimizing asymmetry alone may not explain diverse locomotor phenomena, minimizing a weighted combination of energy and asymmetry, with a small weight on the asymmetry, retains the qualitative predictions of minimizing energy, while providing a further weighting parameter to fit quantitatively to data better (Fig. 2a-b and Extended Fig. 13). We did not seek quantitative fits by tuning parameters in this study; future experiments could delineate the extent to which humans have symmetry as an explicit objective in addition to energy\textsuperscript{63}, given that energy already penalizes large asymmetries\textsuperscript{61}. 
Comparing to minimizing generalized task error

In low-dimensional adaptation tasks such as reaching to a target, the task error to be minimized is unambiguous; for instance, in reaching tasks with visuomotor rotation, the error is defined as angular distance to the reach target.12 However, in higher-dimensional tasks like locomotion, analogous definitions of task error as deviation from desired body kinematics is not uniquely defined. Here, we outline some possible definitions of task error for locomotion and implement them in our modeling framework: for instance, (i) the error could be deviations from desired body states that are walking speed-dependent; further, in this case, the error could be (ii) a weighted sum of the error along different sensory dimensions, with the different dimensions weighted differently. We computed our model predictions when these errors are minimized as the only performance metric (see Methods and Extended Fig. 14a) via the exploration-driven gradient descent of Fig. 1b.

The resulting predictions were not entirely consistent with experiment. We minimized this error for making predictions about split-belt adaptation with different relative weightings of the errors along different dimensions (as in (ii) above). Different relative weightings resulted in distinct behaviors, all of which fell short of fully capturing the experimental findings: the weighting that results in eventual positive step length asymmetry, as seen in experiment, corresponded to energy increase in contrast to experiments, and on the other hand, the weighting that results in monotonic energy decrease has a steady state with substantial negative step length asymmetry, again in contrast to experiments (Extended Fig. 14b-c). A purely kinematic performance metric was similarly found to not explain adaptation to exoskeletons in prior work, where subjects achieved entrainment to exoskeleton impulses or changed their walking frequency without plateating at the unassisted walking kinematics.

Comparison with proprioceptive realignment

Proprioceptive realignment has been proposed as a potential mechanism accounting for the adaptation seen in split-belt locomotion and for reaching tasks with visuomotor perturbations. Vasquez et al. characterized the (proprioceptively) perceived speed of the legs after a split-belt adaptation, effectively finding that humans perceived the fast leg as being systematically slower than reality or the slow leg as faster than reality or both. A causal mechanism relating this sensory recalibration to motor adaptation has not previously been proposed, and a causal mathematical model can help establish whether the sensory recalibration could result in symmetry changes consistent with experiment.

We put forth a mathematical model of proprioceptive recalibration using our framework, which enables linking body dynamics, sensory feedback (both vision and proprioception), and motor action. In our model, the two legs are expected to be on the same surface and proprioceptive deviations from this sensory prediction are perceived as an error to be corrected by recalibrating proprioception, using vision to estimate the perturbation (see Methods and Extended Fig. 12a). This is a type of sensory prediction error, as it is due to a difference between the sensory feedback and what the nervous system expects. This model results in recalibrated estimates of leg speeds such that on a split-belt treadmill, the fast leg feels slower and slow leg feels faster than reality, as in experiment (Extended Fig. 12c), with the recalibration growing in time. The model produces no recalibration when walking on a tied belt, as in experiment. We incorporated this recalibrating proprioceptive sensing as input to the feedback controller without changing other aspects of the default feedback controller to predict what proprioceptive realignment alone can predict.

Proprioceptive realignment as implemented here falls short of explaining qualitative features of split-belt locomotor adaptation. Specifically, while the initial negative step length asymmetry produced by the default controller is decreased by the proprioceptive realignment, the steady state of the adaptation still has substantial negative asymmetry (Extended Fig. 12d), falling substantially short of experimentally observed symmetry and positive step length asymmetry, which is predicted by energy optimization. Interestingly, the model shows coincidental metabolic energy decrease as a result of proprioceptive realignment (Extended Fig. 12e), but this energy decrease is not accompanied by kinematic changes observed in experiment. Thus, while proprioceptive realignment could potentially be a partial cause of split belt adaptation, it does not explain all the associated adaptation phenomena, as also suggested by recent experiments. Beyond split-belt adaptation, proprioceptive realignment cannot explain how humans respond to tied belt speed changes, as experiments did not find significant proprioceptive realignment in the tied belt condition. Finally, proprioceptive realignment via interaction with vision, as implemented here, cannot explain adaptation to purely mechanical changes to the body or the environment such as an added mass or an exoskeleton.
**Extended Data Fig. 12. Sensory recalibration: proprioceptive realignment.**

a) A proprioceptive recalibrator takes in sensory information from proprioceptors of the two legs and vision and computes a recalibrated version of the proprioceptive information. This recalibrated proprioception aligns the proprioception and the vision over a slow timescale.

b) Split-belt protocol.

c) Proprioceptive sensory offset due to sensory realignment.

d) Model predictions: Proprioceptive realignment reduces step length asymmetry but falls short of symmetry.

e) Model predictions: Proprioceptive realignment reduces energy even when energy is not an optimized objective.

Thus, the model predicts that proprioceptive realignment cannot be fully responsible for split-belt adaptation. Proprioceptive recalibration also reduces energy (in this case) even without energy being an explicit objective, but does not reduce energy in other situations such as tied belt adaptation.
**Extended Data Fig. 14. Minimizing task error.** a) We define task error as the deviation from walking kinematics during normal walking at an estimated common belt-speed. b) Split belt protocol. c) Model predictions for minimizing just error without an energy objective (red and yellow); two different weightings are used for the error. Energy minimization is shown for comparison (blue). For the two error weightings shown, one of either step length asymmetry or energy trends disagree with experiment: either the step length asymmetry stops well short of symmetry while decreasing energy somewhat (yellow), or the step length asymmetry goes from negative to positive but energy also decreases and increases again (red).
REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

I am satisfied with the revision and approve of publication.

Reviewer #2 (Remarks to the Author):

I would first like to commend the authors for their thorough responses to the reviewers' concerns. In my opinion, the revised manuscript has been strengthened tremendously, and it is likely to be a landmark study in the field. My remaining concerns are largely requests for clarification of some points that potential readers may misunderstand.

Major Points

While pitting step length asymmetry against energy cost is a clear test of two competing hypotheses, the use of a weighted combination of energy expenditure and step length asymmetry feels like a bit of a heuristic decision to better match simulation results to those from experiments (indeed, the authors allude to this in lines 279 - 281). Is there any theoretical reason to believe that the nervous system cares about both measures? If I understand the proprioceptive realignment results, it seems like a combination of energy optimization and proprioceptive realignment might be a more theoretically sound explanation for the slight over-shoot of symmetry observed experimentally since the Vasquez study provides empirical support for the presence of proprioceptive realignment.

Minor Points

Line 30: I don’t follow the author’s argument that prior studies of adaptation that focus on energy minimization treat locomotion as being episodic. I suggest that a more appropriate assessment is that prior work implicitly assumes that there is some controller that maintains stability while one adjusts their gait pattern to reduce energy cost.

Lines 90 - 91: Please briefly explain what aspects of control are considered to be feedforward. At this point in the manuscript, one might assume that the model relies purely on feedback control.

Line 99: Please briefly explain automatic differentiation for readers who may be unfamiliar with it.

Line 136: Why is Fig. 1a referenced here?

Line 150-151: If I interpret this statement correctly, the authors' model suggests that rapid transients in adaptation are driven by the need for balance control. The authors should consider acknowledging that there is recent empirical evidence to support this idea in the context of split-belt adaptation.
Park S, Finley JM. Manual stabilization reveals a transient role for balance control during locomotor adaptation. J Neurophysiol. 2022 Oct 1;128(4):808-818. doi: 10.1152/jn.00377.2021. Epub 2022 Aug 10. PMID: 35946807; PMCID: PMC9550585.

Buurke TJW, Lamoth CJC, van der Woude LHV, den Otter R. Handrail Holding During Treadmill Walking Reduces Locomotor Learning in Able-Bodied Persons. IEEE Trans Neural Syst Rehabil Eng. 2019 Sep;27(9):1753-1759. doi: 10.1109/TNSRE.2019.2935242. Epub 2019 Aug 14. PMID: 31425041.

Lines 313-315: As it stands, it’s unclear why the authors refer to visual estimates of the perturbation. Why not consider purely proprioceptive estimates? If the authors believe that the addition of vision is necessary, they should consider rewording this sentence to better explain how their simulated recalibration relates to the empirical observations by Vasquez et al.

Figure 2A: Should the y-axis be "metabolic rate" for the model, or is this mechanical power? It would be helpful to describe how these values are quantified and scaled in the caption.

Why is there a drop in model-predicted metabolic rate before the end of the split-belt period?

Reviewer #3 (Remarks to the Author):

I thank the authors for their careful and thorough responses to the initial reviewer comments. I regret that my general impression on the manuscript has changed little since my initial review. This is an impressive undertaking that has undoubtedly required a significant amount of effort. While the authors have attempted to reproduce (and, in many cases, successfully reproduced) a variety of previously published behavioral phenomena, there remains a lack of prospective testing of the model with experiments specifically designed to test model predictions. This is important because several ideas proposed here are controversial (e.g., reinforcement learning as adaptation, shared mechanism underlying exoskeleton and split-belt treadmill walking) and, at times, unsupported by experimental data. As specific examples, noisy split-belt walking does not abolish learning (Exp 3 in Torres-Ovideo and Bastian, J Neurophysiol, 2012, specifically Figs 3 and 5) and one of the anterograde interference predictions in Extended Data Fig 7 has not been demonstrated experimentally (the cited reference by Malone et al provides experimental results for the B-W-A example but not the B-A example; data in support of the B-A prediction are particularly important because studies of reach adaptation do not support these predictions). Other concepts of the model remained unclear to me (e.g., given the simple nature of the bipedal model, it is not clear to me how there can be kinematic changes driven by explicit control that do not affect the energetic estimates that the learning model intends to optimize).
Response to reviewer comments and suggestions.

We thank you all for the kind consideration and for your thoughtful feedback on our manuscript: “Exploration-based learning of a stabilizing controller predicts locomotor adaptation”.

We thank the reviewers for taking the time to consider our revised manuscript. As invited by the editor, we are now submitting a revised manuscript addressing the reviewer concerns. The following provides a point-by-point response to how we have addressed the comments of the reviewers, and directly paste the sections of the paper where changes have been made to address reviewer comments. We quote the reviewer remarks in bold black text and our remarks are in blue regular text.

We are happy to see that reviewers 1 and 2 have been satisfied with our responses, with reviewer 2 providing additional minor comments we have now addressed by revising the writing. Reviewer 3 appreciated the thoroughness with which we had responded to their previous comments and the scale of work involved herein, but suggested new experiments: to address these reviewer 3’s remarks, we have conducted two prospective experiments with 25 human subjects, confirming some model predictions that the reviewer considered critical or surprising, also resolving some seeming contradictions to prior experiments.

We also provide a tracked-changes version of the manuscript per journal policy. In addition to responding to the reviewers, we have edited/tightened the introduction, methods, and figure captions by removing things that are said again later, focusing the details on our central contributions, trying to improve the clarity, but without changing the central takeaways.

Reviewer 1.

I am satisfied with the revision and approve of publication.

Thank you for this remark, and for all your positive and constructive comments in the previous review that have made the manuscript stronger.

Reviewer 2.

I would first like to commend the authors for their thorough responses to the reviewers' concerns. In my opinion, the revised manuscript has been strengthened tremendously, and it is likely to be a landmark study in the field.

Thank you for this remark, and for all your comments in the previous review that have made the manuscript stronger.

My remaining concerns are largely requests for clarification of some points that potential readers may misunderstand.
While pitting step length asymmetry against energy cost is a clear test of two competing hypotheses, the use of a weighted combination of energy expenditure and step length asymmetry feels like a bit of a heuristic decision to better match simulation results to those from experiments (indeed, the authors allude to this in lines 279 - 281). Is there any theoretical reason to believe that the nervous system cares about both measures? If I understand the proprioceptive realignment results, it seems like a combination of energy optimization and proprioceptive realignment might be a more theoretically sound explanation for the slight over-shoot of symmetry observed experimentally since the Vasquez study provides empirical support for the presence of proprioceptive realignment.

We agree with the reviewer’s assessment, and have revised the manuscript to reflect this assessment as follows:

Minor Points

Line 30: I don’t follow the author’s argument that prior studies of adaptation that focus on energy minimization treat locomotion as being episodic. I suggest that a more appropriate assessment is that prior work implicitly assumes that there is some controller that maintains stability while one adjusts their gait pattern to reduce energy cost.

We agree that this was a bit unclear and have replaced the episodic remark by a more specific remark describing what we meant by that (i.e., changes from step to step do not propagate via dynamics) and have revised the relevant sentences as follows:

24 can have short-term and long-term dynamical consequences to stability unless otherwise controlled\textsuperscript{14-17}. Prior accounts of locomotor adaptation\textsuperscript{2,9,18} do not consider the interaction with locomotor dynamics, perhaps assuming that dynamic stability is ensured by a distinct mechanism. For instance, metabolic energy reduction-based accounts\textsuperscript{2,19} treated adaptation to be a univariate optimization process occurring without interaction with the locomotor dynamics – implicitly assuming that changes on one step do not affect the next step through dynamics. Similarly, error-based learning models developed for reaching\textsuperscript{11-13}, when applied to locomotor adaptation\textsuperscript{8,10,18,20,21}, do not interact with the locomotor dynamics; these models fit a single or
Lines 90 - 91: Please briefly explain what aspects of control are considered to be feedforward. At this point in the manuscript, one might assume that the model relies purely on feedback control.

Thank you for this suggestion. We agree that this notion of feedback + feedforward control should be mentioned early. So we have edited the early mention of stabilizing control to note that both feedback and feedforward aspects are involved:

how humans adapt to a novel situation. We characterized this default controller by modeling how humans respond to small deviations from nominal walking on the treadmill\textsuperscript{16,17,35}. The stabilizing controller can be decomposed into a feedforward component, not dependent on the biped’s state, and other state-dependent feedback terms (see Methods). We used the same which precedes the following use of the concept:

Finally, we have replaced ‘stabilizing feedback controller’ with ‘stabilizing controller’ throughout the manuscript when we wished to be inclusive of both feedback and feedforward aspects.

Line 99: Please briefly explain automatic differentiation for readers who may be unfamiliar with it.

We have added the following parenthetical remark.

expanding and revising the earlier briefer sentence below:

Line 136: Why is Fig. 1a referenced here?

Thank you. This was a typo and should have been Fig. 2a. Now fixed!

Line 150-151: If I interpret this statement correctly, the authors' model suggests that rapid transients in adaptation are driven by the need for balance control. The authors should consider acknowledging that there is recent empirical evidence to support this idea in the context of split-belt adaptation.

Park S, Finley JM. Manual stabilization reveals a transient role for balance control during locomotor adaptation. J Neurophysiol. 2022 Oct 1;128(4):808-818. doi: 10.1152/jn.00377.2021. Epub 2022 Aug 10. PMID: 35946807; PMCID: PMC9550585.
Thank you, we have now cited these recent articles, noting that they provide corroborating evidence for our predictions as follows:

Lines 313-315: As it stands, it's unclear why the authors refer to visual estimates of the perturbation. Why not consider purely proprioceptive estimates? If the authors believe that the addition of vision is necessary, they should consider rewording this sentence to better explain how their simulated recalibration relates to the empirical observations by Vasquez et al.

Thank you. We agree that this was unclear. To clarify, only proprioception is re-calibrated to reduce the proprioceptive conflict between the legs — and results in the empirical observations by Vasquez et al. We have clarified the role of vision now as follows:

The old version simply stated the following and was unclear:

 recalibrating proprioception, using vision to estimate the perturbation (see Methods and Extended Fig. 12a). This is a type of

If vision were not used, real differences in the proprioceptive signals from the two legs will get aligned in a manner that the proprioceptive signals are less useful. The constructed model results in no realignment for tied belt changes and realignment for split belt changes, as observed by Vasquez et al.

The Methods section on proprioceptive re-alignment expands on this in the paragraph beginning as follows:

Recalibration takes place when there is substantial conflict between what is expected by the nervous system and what is sensed\textsuperscript{11}. The key missing hypothesis in extending such sensory recalibration to locomotor adaptation lies in the question: what error is the nervous system using to drive recalibration during locomotion? We hypothesize that, given the typical walking experienced in daily life, the nervous system expects the two legs to be on a common surface: this expectation results in a sensory conflict on a split-belt treadmill with both feet experiencing unequal belt speeds.

When the walking surface has fixed speed and the visual environment is uniform, the walking speed can be estimated by the nervous system by two sensory modalities\textsuperscript{10} : vision (based on visual flow) and proprioception (by integrating joint angles and angles rates from muscle spindles and Golgi tendon organs). On a treadmill in a lab, vision has information about how the head moves with respect to the lab, so we identify the visual speed with $v_{\text{body/lab}}$. Proprioception has information about how fast the body parts move relative to the stance foot on the belt, so we identify proprioception with $v_{\text{body/belt}}$. Thus, the body has information to implicitly estimate the belt speed via the following equation: $v_{\text{belt/lab}} = v_{\text{body/lab}} - v_{\text{body/belt}}$. On a split-belt treadmill, all these speeds will be belt-specific, e.g., $v_{\text{belt,1/lab}}$ and $v_{\text{belt,2/lab}}$. The expectation that both legs contact a common surface can be expressed as the equality of these individual belt speeds: $v_{\text{belt,1/lab}} = v_{\text{belt,2/lab}}$. We posit that deviations from this equality result in slow recalibration. Consistent with much of the reaching literature, we recalibrate only the proprioceptive sense, hence the term proprioceptive realignment.

More generally, we edited the Methods section of the proprioceptive alignment section for clarity and additional detail in response to this comment.
Figure 2A: Should the y-axis be "metabolic rate" for the model, or is this mechanical power? It would be helpful to describe how these values are quantified and scaled in the caption.

It is indeed the model's metabolic rate, so this is correct (though the metabolic cost model is based in part on mechanical work). We have now added the following sentence in the caption:

metabolic energy during adaptation and de-adaptation. Metabolic rate is computed as a sum of stance and leg swing cost, and non-dimensionalized (see Methods). b. Walking with an additional mass on one foot². Model qualitatively predicts

Why is there a drop in model-predicted metabolic rate before the end of the split-belt period?

Thank you for catching this. This was a typo in the figure where the split-belt band was mis-sized. We have edited the figure as follows.

Corrected version, where the light green band ends to the right exactly where the 'kink' in the metabolic rate as the reviewer correctly pointed out.

Old version, where the light green band is too wide, as follows:
Reviewer 3.

I thank the authors for their careful and thorough responses to the initial reviewer comments.

Thank you for taking the time to provide detailed comments in your initial review and for your comments below. In your initial review, you had raised numerous interesting challenges to our model -- we are pleased to hear that you found our responses to these challenges thorough.

To recap, in our previous submission, we made the following advancements based on your expert remarks.
- Added an extensive treatment of minimizing step length asymmetry and kinematic task error, showing that they can coincidentally predict some phenomena but not others.
- Advanced a model of proprioceptive conflict, interfacing with locomotor dynamics and stabilizing control, showing that this mechanism can predict some phenomena and not others.
- Acknowledged clearly that the aforementioned adaptation mechanisms can proceed in concert with energy minimization, and that energy minimization is not being posited as the only mechanism.
- Clarified in the discussion section how different dissociable modules within the model can be used to interpret neuroscientific findings.
- While the model we have proposed is about implicit adaptation, we advanced a model of how such implicit adaptation can interact with explicit processes.
- Described more clearly our contributions compared to existing models of locomotor adaptation, which have not been able to explain continuous adaptation while walking without falling.
- Showed how our memory model can capture phenomena that Herzfeld et al 2014 cannot capture.

We are glad these responses thoroughly addressed your comments, except perhaps for your suggestion about prospective experiments, which is now addressed below in response to your next comment.

I regret that my general impression on the manuscript has changed little since my initial review. This is an impressive undertaking that has undoubtedly required a significant amount of effort. While the authors have attempted to reproduce (and, in many cases, successfully reproduced) a
variety of previously published behavioral phenomena, there remains a lack of prospective testing of the model with experiments specifically designed to test model predictions. This is important because several ideas proposed here are controversial (e.g., reinforcement learning as adaptation, …

Our model’s fixed parameters, derived from steady-state non-adaptation data, allowed it to capture unseen phenomena from over fifteen experimental phenomena — which might already be considered a prospective test of the model’s validity. Our goal here was to develop a theoretical model with broad concordance with existing experimental data (as we have demonstrated) to stimulate further model-guided experiments by the research community at large. We considered new experiments beyond the scope of the current manuscript — given, as the reviewer noted, the enormous scale of work already required for this manuscript. Nevertheless, we see the value of prospective experiments, and we have now performed two distinct prospective human subject experiments on a total of 25 subjects, specifically testing model predictions that the reviewer considered surprising or important, and directly resolving the specific concerns the reviewer raised:

**Model prediction-1:** Belt noise can degrade split-belt adaptation as judged by the size of tied-belt after-effects. The reviewer found this result surprising and thought it conflicted with a previous study (Torres-Oviedo and Bastian, 2012).

**Prospective experiment-1:** We performed prospective experiments and showed that belt noise can indeed degrade adaptation.

**Resolution of a seeming contradiction:** We also showed that our model can predict both our new experiment and the old experiment by Torres-Oviedo and Bastian (2012) with the same model parameters, thus resolving any seeming conflict between the two experimental results. Thus, we find that the properties of the noise is important in deciding whether it enhances or degrades learning, and this explains the seeming conflict.

The following modified paragraph is now in the main manuscript, with the newly revised Extended Fig 16 on the next 3 pages.

---

Such degradation of learning due to sensorimotor or device noise is not special to exoskeletons, as we predict that split-belt adaptation can also be degraded via noise-like belt speed fluctuations when the noise is large enough (Extended Data Fig. 16a).

To test this model prediction prospectively, we performed new human subject experiments and compared the post-adaptation after-effects of noise-free and noisy split-belt protocols. We found that subjects had lower after-effects after the noisy adaptation condition, as predicted by the model. Surprisingly, this noise-based degradation is superficially in conflict with earlier experiments by Torres-Oviedo and Bastian, who compared adaptation in a split-belt protocol with gradually increasing speed difference under noise-free and noisy belt speed conditions. They found that the noisy version had higher adaptation as judged by the post-adaptation after-effects, and remarkably, our model captures this improved learning due to noise as well with the same model parameters (Extended Data Fig. 16). Thus, our model is able to explain these superficially conflicting data simultaneously. This example also serves as a caution for deciding the effect of a manipulation based on limited experimental data (e.g., that noise may increase or not degrade learning), because the effect on adaptation of a manipulation may be complex and non-monotonic. Prospective experiments guided by a model such as here can help uncover such phenomena more efficiently.

---

**Model prediction-2:** Our model made specific predictions about interference in the absence of a washout phase, i.e., T-B-A, where B and A are opposing split-belt conditions, and T is a tied-belt condition.

**Prospective experiment-2:** We performed prospective experiments which showed that, as predicted by the model, (1) the initial transient is magnified by experience with a previous counter-perturbation
in the absence of a washout phase, (2) any differences in the time-constant of adaptation is insignificant compared to inter-subject variability.

The new figures and the write-up surrounding these experiments are as follows and the (newly revised Extended Fig 7 follows a couple of pages later.

Contrary to reaching adaptation studies^{37}, our model predicts that anterograde interference need not happen in locomotion: that is, adapting to one perturbation need not make you worse at adapting to the ‘opposite’ perturbation (Extended Data Fig. 7a). This non-interference can be explained by the memory mechanism incorporating a function approximation, so that it can meaningfully extrapolate the learned controllers to the opposite perturbation as well. As predicted by the model, this non-interference was found in prior locomotor experiments^{35}. Because these prior experiments investigating interference had a substantial washout period between the ‘normal’ and the ‘counter’ split-belt speeds^{35}, we performed new prospective experiments without such a washout period; we found that the model correctly predicted the increased initial response due to the counter perturbation and the insignificant changes to adaptation time-constants (Extended Data Fig. 7b).

Figure Redacted
Extended Data Fig. 7. Interference. 'Anterograde interference' refers to the phenomenon where humans sometimes adapt slower to a condition A when they were previously exposed to the 'opposite' condition B, that is, with the belt speed differences reversed between the two belts. a) We performed simulations of two split-belt adaptation protocols: first, T-A-T-A-T-A, alternating between tied-belt conditions T and the split-belt condition A, and second, T-A-T-B-T-A, where one of the A phases is replaced with the opposite condition B. We compare the adaptation between the two protocols during the final A phase (denoted in green), and find that the two protocols are not significantly different in their initial response to the perturbation or the early change in the step length asymmetry, as shown by Malone et al.32. b) In the simulations and experiments, the tied-belt conditions between the split condition served as 'washout'. To test if this non-interference remains in the absence of washout, we performed new experiments in the absence of such a tied-belt washout phase: we compared T-A with T-B-A. We found that the initial step length asymmetry (first step of A) was significantly higher when B was present and the time constant of adaptation during A was not significantly different under the two conditions (paired t-test, t(df) = 4.36, p < 0.05). This confirmed our model simulations which predicted that the initial transients will be higher after B. The model also predicted no statistically significant difference in the adaptation time constant in the presence of same subject number and inter-subject variability of magnitude similar to that in the experiment. Box-plot shows median, 25-75% percentile and range; p values computed showing insignificant differences are from t-tests had p > 0.3.
The Methods section for the two prospective experiments is as follows:

**Prospective experiments**

The computational model we put forth here can be used to design prospective experiments, augmenting experimenter intuition. Here, we conducted two model-guided experiments to test predictions of the model that are surprising when compared to the existing literature: (1) on the effect of environment noise on locomotor adaptation, and (2) on the effect of an immediately preceding counter-perturbation on a subsequent adaptation.

Twenty five subjects (16 male, 9 female, self-reported sex, age 21.9 ± 3 years, mean ± s.d.) participated with informed consent and the experiments were approved by the Ohio State University IRB. Subjects were assigned randomly into two groups: sixteen subjects performed experiment 1 (which contained two conditions with eight subjects in each), and nine subjects performed experiment 2. Both experiments involved walking on a split-belt treadmill (Bertec Inc.), with the details of the protocol provided below. Foot movement was tracked via a Vicon T20 motion capture system.

Experiment 1 was designed to test the model prediction that when the belt noise level was sufficiently high, learning can be degraded, which is surprising relative to a prior finding that a modest level of belt noise can slightly enhance learning as measured by after-effects\(^{50}\). For this experiment, the subjects were sub-divided into two groups of eight subjects: one group performed a no-noise abrupt protocol (Extended Figure 8), in which subjects started walking under tied-belt conditions at 0.9 m/s, then adapted to split-belt condition of 0.6 m/s and 1.2 m/s kept constant for 10 minutes, followed by three minutes of tied-belt walking at 0.6 m/s; the second group had an identical protocol except the split-belt condition involved continuously changing belt speed for just the fast belt, fluctuating in a piecewise linear manner with zero mean and 0.2 m/s standard deviation (normally distributed). The consecutive grid points of the piecewise linear noise were separated by 1.2 seconds, roughly equal to a stride period, so that the noise value was different two strides apart (the noise in\(^{50}\) was changed every 3 seconds, and thus had greater temporal correlation); speed changes had 0.1-0.2 m/s\(^2\) accelerations. The post-adaptation after-effect in step length asymmetry after baseline subtraction, averaged over the first 8 strides (about 10 seconds) was used as a measure of adaptation similar to prior work\(^{50}\). We compared these after-effects between the noise and no-noise case, testing the hypothesis that the noise case has lower after-effects.

Experiment 2 was designed to test the model prediction regarding savings, specifically whether experiencing a counter perturbation B beforehand, interferes with adaptation to perturbation A. Previous experiment had found that if B and A were separated by a washout period, the adaptation to A was not significantly affected compared to not having experienced B. Our model had a distinct prediction for when A immediately followed B, without a washout period W. So, subjects performed this experimental protocol T-B-A (Figure 7) in which 4 minutes of walking on a tied-belt at 0.9 m/s (T) was followed by a split-belt condition with belt speeds of 0.6 and 1.2 m/s for 10 minutes (B), immediately followed by the opposite split-belt condition 1.2 and 0.6 m/s for 10 minutes. Equivalent to comparing the A of protocols T-A and T-B-A by symmetry, we compared the initial transient and the time-constant of the two adaptation periods B and A in T-B-A: the B of T-B-A was without prior split-belt experience and the A of T-B-A is the adaptation phase just after a counter-perturbation.

… shared mechanism underlying exoskeleton and split-belt treadmill walking), unsupported by experimental data. As specific examples, noisy split-belt walking does not abolish learning (Exp 3 in Torres-Ovideo and Bastian, J Neurophysiol, 2012, specifically Figs 3 and 5).

As noted in response to the previous comment, we have resolved this seeming conflict via prospective experiments. Our model can simultaneously predict enhancement of learning by noise as found in Torres-Ovideo and Bastian, J Neurophysiol, 2012 and find experimental degradation of learning by noise when the noise has appropriate properties, as predicted by our model.

We were intrigued that the reviewer felt that the mechanisms underlying exoskeleton and split-belt adaptation may be fundamentally different, given that it is the same human motor control system interacting with both devices, the task is similar (walking), and the devices also have some core similarity. To address this, we have expanded the following sentences in the Discussion:
Our model suggests explanations for why humans may adapt reliably in some novel situations (for instance, during split-belt walking, and not others (for instance, some exoskeleton studies). One might wonder if common principles underlie reliable adaptation in one class of devices and unreliable adaptation in another class, given that both devices are interacting with the same human motor control system. First, we note that both exoskeletons and split-belt treadmills share a core dynamical similarity: they are both mechanical devices that contact the body, applying forces and performing positive or negative work of different specifications. Second, our results also suggest that the two classes of devices are not fundamentally different with respect to motor learning, but that the differences may be due to dynamical properties of treadmills versus current exoskeletons and their controllers. Our model suggests ways in which we can make subjects less reliable learners on split-belt treadmills and reliable learners on exoskeletons and prostheses. Specifically, our model predicts that split-belt adaptation can be degraded by

\[ \text{explicit control} \]

... and one of the anterograde interference predictions in Extended Data Fig 7 has not been demonstrated experimentally (the cited reference by Malone et al provides experimental results for the B-W-A example but not the B-A example; data in support of the B-A prediction are particularly important because studies of reach adaptation do not support these predictions).

Thank you for the suggestion to test our model against B-A protocols, without washout. As noted in response to a previous comments, we have now performed these prospective experiments and found that our model satisfactorily captured the key aspects of the adaptation response (as described in response to the earlier comment).

Other concepts of the model remained unclear to me (e.g., given the simple nature of the bipedal model, it is not clear to me how there can be kinematic changes driven by explicit control that do not affect the energetic estimates that the learning model intends to optimize).

We assume the reviewer is referring to the model of how explicit control and implicit adaptation interact. The kinematic changes driven by explicit control do affect energy cost. We suspect that the confusion may be due to how kinematic changes due to explicit control are retained when the explicit control is removed even though they have effects on energy cost. They are not retained because the output of the explicit and implicit pathways are in parallel and additive (as in the Jordan-Taylor model architecture), progress made by the explicit pathway is lost when the explicit pathway is turned off.

We also note that how the kinematic changes affect the energy cost within the biped model has been empirically validated via extensive past biomechanics experiments (Donelan, Kram, and Kuo 2002, Doke and Kuo 2005), as noted in greater detail in the Supplementary Text.
REVIEWERS' COMMENTS

Reviewer #2 (Remarks to the Author):

The authors have adequately addressed all of my concerns. I look forward to seeing the impact that this work will have on the field.

Reviewer #3 (Remarks to the Author):

I thank the authors for their significant efforts in revising their work; no additional comments.