1. Make clear our goals in letter and Intro [Is there contribution here?]; be clear that primary purpose is to illustrate the framework and to show consequences; it is not trivial problem, we need to break the project into multiple parts; we didn’t know until recently these findings are so ambiguous in distinguishing among these theories.
   1. Shift the burden of proof
   2. The computational framework we provide, we believe, provides an important role in addressing the noted issue.
2. Intro: The solution is not simple. It is more complicated than we here presented. We only compared the models in qualitative terms, it is not the models that are complex. [Aim to get the reviewers on board with us.]
   1. all of this we have said so far tells us that there might be an issue; but the model results will tell us whether there is indeed a possibility to distinguish between the accounts or not based on existing empirical data.
   2. [manage the expectation] To anticipate, the take-home point is that there is a substantial problem we need to pay attention to. While the framework we present here provides a great starting point to a possible solution, this paper does not. Ideally we should know the solution already, but we do not. In this situation, we still think that it is still of prime importance to make clear this problem. In this paper, we are not yet presenting simulations that can determine the best testing scenarios [in letter: it is not a trivial issue; that is a major goal of a multi-year NSF grant that we submitted; USE the language to indicate the feasibility, but also beyond the scope of this paper], but we provide general points that can guide future experimental design.
3. Add a section in the GD about other types of effects in speech adaptation which may be explained by the current framework;
   1. One concern readers have may be that we have only shown two phenomena. Should we be worried about this. Are there other results that may be accounted for by the three mechanisms?
   2. [Is there any data available that should be accounted for by the current framework?] To us, it is not apparent what type of data that would be. The reason is that there has been simply no direct contrasts in experiments. To some extent, there are some notable exceptions. For instance, work from Lori Holt (summarize) tells us that there are really short-term changes that are not intuitively understood as representational learning. That leaves open whether normalization is also sufficient to account for adaptation for, for instance, foreign accents. These findings do not address the question we are asking and the question we need to ask. To us, the important issue is that we need to move beyond the current way of testing in order to get decisive empirical data.
   3. [how likely are our results generalizable?] While there are many different paradigms, incl. the ones we grouped together, they actually differ greatly in terms of the tasks, stimuli, etc.
   4. For what’s worth, we conducted another simulation but opted to not present it here due to the length of the paper. [all boils down to the fact that the qualitative data from existing work that does not allow us to distinguish between these accounts] It supports the general idea that any of the three mechanisms may work to account for something like PR, AA – in the literature, the common perception is that these are different phenomena pursued in different lines of work. [say in the letter that this point is not self-evident]

**To-do**

1. Divide the praise, problems that we can address, misunderstanding of our goals/scope.
2. Go through to see we can move a large chunk to Appendix. **Color mark the paper and indicate what could go, what are essential, what can be condensed**
   1. – What if Sec2 is gone? How accessible is Sec3?
   2. If we keep Sec2 largely, then we need to emphasize in the Introduction that this is the first time that any formal framework is applied, it is necessary in order to for us to understand what the problem is, which is a prerequisite for finding a solution.
3. Write down our take-home points; and what we want to emphasize in the conclusion section – a) brief but clear; b) echo what we said in the paper
4. Discussion on Neural findings:
   1. p.9 – reduce the level of details in the introduction and anticipate that we will relate our framework to neural studies.

Dear Dr. Xin Xie,  
  
Thank you for submitting your work to Cortex. Your above paper has now been reviewed by expert referees, whose comments are enclosed for your perusal. On the basis of these comments, we cannot accept your manuscript in its present form but would like to invite you to revise your paper, taking into account the issues raised by the reviewers. Please note that acceptance is not guaranteed at this stage and any revision is likely to be sent back to the referees for further review.  
  
Both reviewers offered many positive comments regarding the paper, however, a few concerns need to be addressed if you choose to submit a revision. Here, I highlight four important points but the authors should include a point-by-point response to each of the reviewer comments, highlighting each change made in your manuscript and/or providing a suitable rebuttal.  
  
1)      Reviewer 1 suggests offering more insight into whether there exist conditions for which any of the proposed mechanisms can be ruled out or whether it may be more about finding the most optimal solutions. Illustrating such an example with a simple case might add some clarification to the main conclusions.  
  
2)      Reviewer 2 commented on the discussion related to neural effects. Here, the authors can focus on offering a more explicit and clear link on the relationship between the proposed neural mechanisms and the computational mechanisms, rather than expanding the length of this section. **Highlighting the main mechanistic arguments made based on specific neural data and drawing parallels to the computational models (or if appropriate, pointing to any neural evidence that distinguishes among or brings together different computational possibilities would be useful).** This section will provide a nice connection to other topics addressed in the special issue.  
  
3)      Both reviewers commented on the length of the manuscript. Perhaps some of the mathematical details of the models could be described in detail in an appendix instead? I think this will increase the impact of the main points made for those who are not computationally-inclined. However, it is important to have them readily available in an appendix for those who would like to know more about the specifics and default parameter assumptions.  
  
4)   Finally, both reviewers commented on the need for summarizing the key take-home messages and more specific steps for moving forward, in the conclusions section.

Reviewer #1: Review of : What we do (not) know about mechanisms underlying adaptive speech perception: A computational review  
  
SUMMARY  
The authors introduce a computational framework that characterizes how recent exposure can impact pre-linguistic signal normalization, linguistic representations, and decision-making.  They use this model to raise questions about whether prior accounts of adaptive change necessarily relate to changes in linguistic representations, or whether they could be attributable to other levels of representation, across two case studies.  They conclude by offering targeted guidance for how future work could leverage this framework to improve experimental practice and better determine the locus (or loci) of adaptive change in speech perception.  
  
ASSESSMENT  
Overall, I thought this was a very strong paper.  The computational framework introduced in this work and illustrated in the two case studies makes a strong argument that attributing adaptive change to changes in linguistic representations is premature --- both pre- and post-linguistic adaptation can also potentially explain these results.   Incorrectly oversimplifying the potential contributions of other cognitive systems (e.g., the decision system when studying the lexical system, or vice versa) clearly has the potential for misattributing the locus of a range of empirical effects, but direct comparisons of the potential contributions of each system are lacking in the literature, which this work aims to address.

This work also very clearly demonstrates the utility of models not only in describing extant empirical results but in generating targeted predictions that can delineate between competing (or partially complementary) theoretical accounts, which seems especially useful in the present case.

I also appreciate how the current framework can be extended to capture an even broader set of phenomena, which speaks to the generality of the approach.

Furthermore, the introduction of a formal, quantitative method appears to have had the added benefit of revealing how, sometimes by chance, a set of stimuli could have more variation on a dimension other than the typically expected one, and how that could alter the results of the experiment (in section 4.2.1).  On another front, the authors went to great lengths to integrate this paper with a wide body of prior empirical and theoretical work.  Although this does make for a long read, it does help illustrate the breadth of issues that this work connects to as it is presented and could be connected through in future extensions.  I have gone back and forth on suggesting that the length of the paper be reduced, which might increase its uptake, but I do think that there is value in having all of this content available in one place.

The authors' commitment to open-science principles is also truly exemplary, in that they provide literally all materials needed to replicate their work, including the paper itself, in their materials.  In so doing, they have increased the accessibility of this computational approach to other researchers and made it easier to test other related hypotheses or model other related data sets in the future.  I found the extensive set of figures very helpful in fleshing out a number of points and in illustrating how various model parameters operate, and the model description to be reasonably accessible, although the complexity of some aspects of the math may still pose some accessibility challenges for a wide audience.   
  
Notwithstanding these strengths, there were a few points that I thought might warrant some revision.  In no particular order:

I found the section on neural correlates to be somewhat underwhelming and non-specific**.  For this paper to have substantial weight, I think it would be necessary to spell out how specific neural measures could be tied, in a quantitative way, to specific parameters of the computational framework.**  I would encourage the authors to consider a simple case study to illustrate this point, otherwise, **I would suggest that this section could be tightened considerably or eliminated**.

In a different vein, the current case studies attempt to illustrate the separate and independent effects of manipulating parameters at three different levels of processing; however, I would imagine that there is a good case to argue that adaptation could occur at all three levels simultaneously, at least to some degree.  If that is the case, I think it would be useful to explain how change could be modeled at all three levels simultaneously, even if the main focus here is on these clear contrasts that are possible through independent manipulations of each parameter.

Finally, I found that a clear integrative conclusion was missing for this paper, which seemed to end fairly abruptly.  I think it would be very helpful for a reader of a 60+ page paper to have the authors state what they think the key take-home points of the work are to facilitate retention of those points and to avoid losing them among the dense package of content that this paper provides.

Other comments:  
  
  
-when introducing the lapse rate parameter on p. 21, I was not initially sure of why this parameter would receive such prominent treatment in the paper, although the case was nicely made later on in the paper.  Given the importance of lapses was not discussed in detail earlier in the paper, it could be helpful to foreshadow the importance of this parameter earlier on.   
  
-I had several issues using the pdf document, including generating a printed copy.  I suggest the journal and the authors be mindful of this if this paper is moved to production.  I was on windows 10 using the current version of Adobe Acrobat when I encountered these issues.   
  
-likewise, a number of the figures do not print out well, potentially because of the default state of the figures in the interactive figures.  Although I appreciate that the figures are best appreciated on a computer, it would be useful, I think, if they could work in a basic sense in a printed copy of the document.   
  
-on p.36: Examining Figure 14A, my impression was that different stimuli were used in the /d/-shifted vs. /t/-shifted panels of the figure.  Would the tightest control not contain the same base stimuli shifted in either direction?   
  
p. 37: The focus on the simulations is on the beginning of the test phase; however, should the model not also be able to account for performance throughout the test phase?  If not, why not?  Is this reflective of some additional parameter not included in the model (e.g., a reluctance to keep changing beyond a certain point?), particularly in the face of repeated stimuli?  
  
p.55 the authors state "the highest accuracy is obtained for the fastest changes, and it matches that observed for changes in decision making."  Looking at the data, I am not sure that the match is especially strong, but I may be misinterpreting the data being referenced here or the level of "match" that the authors are referring to.  Perhaps this could be clarified?   
  
  
Reviewer #2: Summary  
The manuscript presents a model of phonetic adaptation effects that aims to better investigate the linking functions between behavioral tasks and the mechanisms of speech adaptation. In particular, the model distinguishes between processes of altering representation, establishing response biases and normalizing input. The authors then simulate typical talker adaptation and accent adaptation effects, exploring the parameter space of the model for these three different processes. They find that certain regions of parameter space can recreate the qualitative patterns of these tasks for all three types of process. They interpret this as a major challenge to prominent accounts of talker and accent adaptation, which regularly posit a locus of effect in altered representations. They then suggest that use of their approach to simulating behavioral adaptation tasks can help better identify the mechanisms of adaptation, particularly when paired with careful stimulus and design decisions to maximize how well patterns of data can discriminate between the different processes.  
  
Review  
There's a lot to like about this manuscript. The general approach is timely and important - the authors emphasize the need to carefully think about the way we operationalize the constructs we care about, and provide a mathematical model to do so. I applaud them for this approach - many of our theories are built on a backbone of methods that haven't received this kind of careful methodological treatment, and instead rely heavily on face validity without adequate skepticism. Assessing the validity of our operationalizations is an important next step to overcome to the replication crisis and strengthen our theories.  
  
The paper is also written extremely clearly. The model is described well, the theoretical backing is easy to follow, and the authors show a very comprehensive mastery of the relevant literature.  
  
Despite these clear pros, the paper is a bit challenging to review. It's long, thorough and detailed, but I'm left at the end wondering what, exactly, we've gained. The authors nicely identify that the interpretations given for past studies could arise because of different mechanisms, but I'm not sure that a 90-page technical modeling paper is necessary for this. Much of the introduction highlights the theoretical reasons that different mechanisms might explain the extant behavioral data. The fact that the model confirms this is reassuring of their logic, but also makes me wonder what we learn from the model itself. **The primary surprising finding from the model seemed to be that response biases can do more than we might expect, but this feels a little underwhelming given the scale of the paper.**  
  
In the General Discussion, the authors suggest ways to use this approach to better investigate mechanisms of adaptation, but this felt fairly underspecified. For example, I agree that quantitative model fit might prove more important than qualitative fit in some cases, but relying on quantitative fit is easier said than done. In the case of the present model, **there are multiple parameters for each mechanism that can be manipulated in search of best fitting models, and it's not clear what the best way to adjudicate between these might be**. If the normalization mechanism leads to a slightly better quantitative fit than the representation mechanism, but does so only in a very small region of parameter space, should we count this as evidence that it's a more likely candidate mechanism?

**This issues of model selection on the basis of quantitative fit are quite contentious, and seem particularly challenging for complex models with numerous adjustable parameters, like the present one.** In addition, the way the model accommodates training input depends on how it establishes and represents its priors about category information. Some aspects of this are included as adjustable parameters in this model, but others aren't - for example, the way that the model maps acoustic input into perceptual features relies on assumptions about how input is processed and normalized. It's necessary to make some assumptions like these, but they then pose challenges to interpretations that rely on quantitative fit comparisons. Might a different parameter set have proven better under different assumptions of how input maps to perceptual representations?  
  
This concern points to a benefit for situations where predictions of qualitative distinctions are meaningful. **In particular, it would be helpful if the authors could identify conditions that can't be accounted for by some of the mechanisms, no matter the parameter choice.** For example, are there certain types of stimuli or training regimens that would only predict an effect if representations change, but can't be explained by normalization or response bias? **Can the authors point to any truly discriminant measures by which we can rule out a mechanism as incapable of explaining a pattern of results, rather than just offering a poorer quantitative fit?** Or is the whole enterprise here **a question of finding the specific region of parameter space that best accommodates whatever data can be collected?** This isn't necessarily disqualifying - ideas like parameter space partitioning have proven a useful tool for comparing simulations - but it raises questions about whether the model is just overly flexible. **Can we fit basically all the same qualitative patterns of data with each mechanism, if we find the right parameterizations?** And if so, is it worth doing a more formal parameter space partitioning analysis to see if some of the approaches more stably predict this?  
  
There's a wide array of other speech adaptation and/or talker normalization tasks out there beyond those simulated here. Are there any of these for which the model offers qualitatively discriminant predictions?  
  
Overall, I think the approach taken in this paper is intriguing and potentially valuable, but it feels like its **missing the specificity needed to explain next steps.** The ability of the model to simulate patterns of data from all three mechanisms is problematic for previous accounts - so what should we do? The paper would be much more powerful if it had a clearer path forward to help us discriminate between these mechanisms. There's a short paragraph on Page 62 that starts down this path, but it quickly reverses into more conceptual discussion of how this could be done.  
  
I find I'm left wanting more of something from this paper, despite it already being lengthy. I love the perspective about a need for more careful consideration of linking functions, but think that much of the heavy lifting for this perspective can be carried by the introduction and lit review, before the model has been formally described, let alone the simulations have been run. I like the premise of using this modeling approach to adjudicate between different mechanisms of adaptation, but didn't feel like we got a clear answer of how to do so with this model. We certainly need a better understanding of the link between behavior and adaptation tasks and theories of the processes that allow this. The issue is that this paper just points out what's wrong, without enough thought for how to make it right.