Official Response

Dear Dr Schapiro and Prof. Frank,

We would like to thank the reviewers for their thorough, constructive, and positive appraisal of our manuscript titled *"Rapid learning of predictive maps with STDP and theta phase precession"*. In response to the points raised, we have made numerous improvements to the manuscript to enhance the links to existing literature, demonstrate the biological plausibility of the model and provide theoretical insight into how and why it works. As such, we have revised the original manuscript, taking care to present a clear message suitable for *eLife's* broad audience. We now present that manuscript for re-submission, with new and changed text coloured blue. The changes made to accommodate the essential revisions include:

1) Significantly more discussion of the work's relationship to relevant prior models of the hippocampus (as described by Reviewer #1)

We have added a large quantity of text addressing the work's relationship to relevant prior models of the hippocampus. We have added substantially to the introduction and discussion, and also have made other additions throughout the results to provide better context.

2) New simulations that address Reviewer 2's concerns about biological plausibility.

We have performed several new simulations, producing new results that speak to the model's robustness and biological plausibility, constituting 3 entirely new multipanel supplementary figures examining the effects on the model of place field size, running speed, phase precession parameters, weight initialisation, weight update regimes and downstream phase precession in CA1.

3) Analysis that sheds light on why theta sequences + STDP approximates the TD algorithm (as described by Reviewer #2).

A significant new theoretical section provides mathematical insight as to why a combination of STDP and theta phase precession can approximate the temporal difference learning algorithm.

We again thank the editors and the reviewers for the thoughtful and constructive review. Below we have provided a detailed point-by-point response to each reviewer with their initial reviews indented and in *italics*, followed by our response, and excerpts from the updated manuscript in "*quotations*" and small font with changes coloured in blue.

Reviewer #1 (Public Review):

> *The authors focused on linking physiological data on theta phase precession and spike-timing-dependent plasticity to the more abstract successor representation used in reinforcement learning models of spatial behavior. The model is presented clearly and effectively shows biological mechanisms for learning the successor representation. Thus, it provides an important step toward developing mathematical models that can be used to understand the function of neural circuits for guiding spatial memory behavior.*
>
> *However, as often happens in the Reinforcement Learning (RL) literature, there is a lack of attention to non-RL models, even though these might be more effective at modeling both hippocampal physiology and its role in behavior. There should be some discussion of the relationship to these other models, without assuming that the successor representation is the only way to model the role of the hippocampus in guiding spatial memory function.*

We thank the reviewer for the positive comments about the work, and for the detailed and constructive feedback. We agree with the reviewer that the manuscript will benefit from significantly more discussion of non-RL models, and we've detailed below a number of modifications to the manuscript to better incorporate prior work from the hippocampal literature, including the citations the reviewer has listed. Since our goal with this paper is to contextualise hippocampal phenomena in the context of an RL learning rule, this is really important and we appreciate the reviewers recommendations. We have added text (outlined in the point-by-point responses below) to the introduction and to the discussion that we hope better demonstrates the connections between the SR and existing computational models of hippocampus, and communicates clearly that the SR is not *unique* in capturing phenomena such as factorization of space and reward or capturing sequence statistics, but is rather a model that captures these phenomena while also connecting with downstream RL computations. Existing RL accounts of hippocampal representation often do not connect with known properties of hippocampus (as illustrated by the fact that TD learning was proposed in prior work to be the learning mechanism for SRs, even though this doesn't have an obvious mechanism in HPC), so the purpose of this work is to explore the extent to which TD learning effectively overlaps with the well-studied properties of STDP and theta oscillations. In that sense, this paper is an effort to connect RL models of hippocampus to more physiologically plausible mechanisms rather than an attempt to model phenomena that the existing computational hippocampus literature could not capture.

> *1. Page 1- "coincides with the time window of STDP" - This model shows effectively how theta phase precession allows spikes to fall within the window of spike-timing-dependent synaptic plasticity to form successor representations. However, this combination of precession and STDP has been used in many previous models to allow the storage of sequences useful for guiding behavior (e.g. Jensen and Lisman, Learning and Memory, 1996; Koene, Gorchetchnikov, Cannon, Hasselmo, Neural Networks, 2003). These previous models should be cited here as earlier models using STDP and phase precession to store sequences. They should discuss in terms of what is the advantage of an RL successor representation versus the types of associative sequence coding in these previous models.*

We agree that the idea of using theta precession to compress sequences onto the timescale of synaptic learning is a long-standing concept in sequence learning, and that we need to be careful to communicate what the advantages are of considering this in the RL context. We have added these citations to the introduction:

"One of the consequences of phase precession is that correlates of behaviour, such as position in space, are compressed onto the timescale of a single theta cycle and thus coincide with the time-window of STDP O(20 − 50 ms) [8, 18, 20, 21]. This combination of theta sweeps and STDP has been applied to model a wide range of sequence learning tasks [22, 23, 24], and as such, potentially provides an efficient mechanism to learn from an animal's

experience – forming associations between cells which are separated by behavioural timescales much larger than that of STDP."

and added a paragraph to the discussion as well that makes this clear:

"That the predictive skew of place fields can be accomplished with a STDP-type learning rule is a long-standing hypothesis; in fact, the authors that originally reported this effect also proposed a STDP-type mechanism for learning these fields [18, 20]. Similarly, the possible accelerating effect of theta phase precession on sequence learning has also been described in a number of previous works [22, 55, 23, 24]. Until recently [40, 41], SR models have largely not connected with this literature: they either remain agnostic to the learning rule or assume temporal difference learning (which has been well-mapped onto striatal mechanisms [37, 56], but it is unclear how this is implemented in hippocampus) [54, 31, 36, 57, 58]. Thus, one contribution of this paper is to quantitatively and qualitatively compare theta-augmented STDP to temporal difference learning, and demonstrate where these functionally overlap. This explicit link permits some insights about the physiology, such as the observation that the biologically observed parameters for phase precession and STDP resemble those that are optimal for learning the SR (Fig 3), and that the topographic organisation of place cell sizes is useful for learning representations over multiple discount timescales (Fig 4). It also permits some insights for RL, such as that the approximate SR learned with theta-augmented STDP, while provably theoretically different from TD (Section 5.8), is sufficient to capture key qualitative phenomena."

> *2. On this same point, in the introduction, the successor representation is presented as a model that forms representations of space independent of reward. However, this independence of spatial associations and reward has been a feature of most hippocampal models, that then guide behavior based on interactions between a reward representation and the spatial representation (e.g. Redish and Touretzky, Neural Comp. 1998; Burgess, Donnett, Jeffery, O'Keefe, Phil Trans, 1997; Koene et al. Neural Networks 2003; Hasselmo and Eichenbaum, Neural Networks 2005; Erdem and Hasselmo, Eur. J. Neurosci. 2012). The successor representation should not be presented as if it is the only model that ever separated spatial representations and reward. There should be some discussion of what (if any) advantages the successor representation has over these other modeling frameworks (other than connecting to a large body of RL researchers who never read about non-RL hippocampal models). To my knowledge, the successor representation has not been explicitly tested on all the behaviors addressed in these earlier models.*

We agree – a long-standing property of computational models in the hippocampal literature is a factorization of spatial and reward representations, and we have edited the text of the paper to make it clear that this is not a unique contribution of the SR. We have modified our description of the SR to better place it in the context of existing theories about hippocampal contributions to the factorised representations of space and goals, and included all citations mentioned here by adding the following text.

We have added a sentence to the introduction:

"However, the computation of expected reward can be decomposed into two components – the successor representation, a predictive map capturing the expected location of the agent discounted into the future, and the expected reward associated with each state [26]. Such segregation yields several advantages since information about available transitions can be learnt independently of rewards and thus changes in the locations of rewards do not require the value of all states to be re-learnt. This recapitulates a number of long-standing theories of hippocampus which state that hippocampus provides spatial representations that are independent of the animal's particular goal and support goal-directed spatial navigation[27, 28, 23, 29, 30]"

We have also added a paragraph to the discussion:

"The SR model has a number of connections to other models from the computational hippocampus literature that bear on the interpretation of these results. A long-standing property of computational models in the hippocampal literature is a factorisation of spatial and reward representations [27, 28, 23, 29, 30], which permits spatial navigation to rapidly adapt to changing goal locations. Even in RL, the SR is also not unique in factorising spatial and reward

representations, as purely model-based approaches do this too [26, 25, 67]. The SR occupies a much more narrow niche, which is factorising reward from spatial representations while caching long-term occupancy predictions [26, 68]. Thus, it may be possible to retain some of the flexibility of model-based approaches while retaining the rapid computation of model-free learning."

> *3. Related to this, successes of the successor representation are presented as showing the backward expansion of place cells. But this was modeled at the start by Mehta and colleagues using STDP-type mechanisms during sequence encoding, so why was the successor representation necessary for that? I don't want to turn this into a review paper comparing hippocampal models, but the body of previous models of the role of the hippocampus in behavior warrants at least a paragraph in each of the introduction and discussion sections. In particular, it should not be somehow assumed that the successor representation is the best model, but instead, there should be some comparison with other models and discussion about whether the successor representation resembles or differs from those earlier models.*

We agree this was not clear. This is a nuanced point that warrants substantial discussion, and we have added a paragraph to the discussion (see the paragraph in the response to point 1 that begins *"That the predictive skew of place fields can be accomplished…*").

> *4. The text seems to interchangeably use the term "successor representation" and "TD trained network" but I think it would be more accurate to contrast the new STDP trained network with a network trained by Temporal Difference learning because one could argue that both of them are creating a successor representation.*

We now refer to these as "STDP successor features" and "TD successor features". We have also replaced all references of "true successor representation/features" to "TD successor representation/feature" and have edited the text at the beginning of the results section to reflect this:

"The STDP synaptic weight matrix W_{ij} (Fig. 1d) can then be directly compared to the temporal difference (TD) successor matrix M_{ij} (Fig. 1e), learnt via TD learning on the CA3 basis features (the full learning rule is derived in Methods and shown in Eqn. 27). Further, the TD successor matrix M_{ij} can also be used to generate the 'TD successor features'..."

Reviewer #1 (Recommendations for the authors):

> *Page 4 - top line - "in the successor representation this is because CA3 place cells to the left...". I think this is confusing as the STDP model essentially generates the same effect. I think this should say: "In the network trained by Temporal Difference learning this is because CA3 place cells to the left...". This better description is used further down where the text says "between STDP and TD weight matrices". Throughout the manuscript*

Thank you for this suggestion. We've gone through the text and implemented this change where the issue arises, as well as adding the sentence clarifying our terms (described in the in response to the public review in response to point 4).

> *Page 4 - end of the first paragraph - "potentially becoming negative" - it is disconcerting to have this discussion of the idea of synaptic weights going from positive to negative in the context of the STDP model. One of the main advantages of this model is its biological realism, so it should not so casually mention violating Dale's law and having the synapse magically switch from being glutamatergic to GABAergic. This is disturbing to a neuroscientist.*

Thank you for this valid point – we've added the following line to follow that sentence:

"So, for example, if a postsynaptic neuron reliably precedes its presynaptic cell on the track, the corresponding weight will be reduced, potentially becoming negative. We note that weights changing their sign is not biologically plausible, as it is a violation of Dale's Law [43]. This could perhaps be corrected with the addition of global excitation or by recruiting inhibitory interneurons."

> *Page 4- "is an essential element of this process." - The importance of theta phase precession to sequence learning with STDP has been discussed in numerous previous papers. For example, in a series of four papers in 1996, Jensen and Lisman describe in great detail a buffer mechanism for generating theta phase precession, and show how this allows encoding of a sequence. This is also explicitly discussed in Koene, Gorchetnikov, Cannon, and Hasselmo, Neural Networks, 2003, in terms of a spiking window of LTP less than 40 msec that requires a short-term memory buffer to allow spiking within this window.*

We agree that the paper would benefit from better connection with the prior work on sequence learning with STDP and have added text to the introduction and discussion. In the introduction, we have added:

"One of the consequences of phase precession is that correlates of behaviour, such as position in space, are compressed onto the timescale of a single theta cycle and thus coincide with the time-window of STDP O(20 − 50 ms) [8, 18, 20, 21]. This combination of theta sweeps and STDP has been applied to model a wide range of sequence learning [22, 23, 24], and as such, potentially provides an efficient mechanism to learn from an animal's experience – forming associations between cells which are separated by behavioural timescales much larger than that of STDP."

And we've included a paragraph to the discussion to make this clear. This is contained in the paragraph above, in our response to point 1 in the public review (see paragraph starting *"That the predictive skew of place fields can be accomplished…*").

> *Page 4 - "our model and the successor representation" - again this is confusing and should instead contrast "our model and the TD trained successor representation"*

Thank you, we have made this change to the text.

> *Page 6 - "in observed" - is observed.*

Thank you - fixed.

> *Page 6 - "binding across the different sizes" - This needs to be stated more clearly in the text as it is very vague. I would suggest adding the phrase: "regardless of the scale difference".*

Thank you for the suggestion – we have implemented this change.

> *Fig. 4D - "create a physical barrier" - this is very ambiguous as it recalls a physical barrier in the environment as between two rooms - should instead say "created an anatomical segregation".*

Thank you for the suggestion – we have implemented this change.

> *Page 8 - "hallmarks of successor representations" - there should be citations for what paper shows these hallmarks of the successor representation.*

Thank you – we have added citations to Stachenfeld et al 2014, Stachenfeld et al 2017, and de Cothi & Barry 2020 to this sentence.

Page 8 - "arrive in the order" - Here is a location where citations to previous papers on the use of a phase precession buffer to correctly time spiking for STDP should be added (i.e. Jensen and Lisman, 1996; Koene et al. 2003).

Thank you for the suggestion – we have implemented this change.

Page 8 - "via Hebbian learning alone" - add "without theta phase precession" to be clear about what is not being included (since it could be anything such as other aspects of a learning rule).

Thank you for the suggestion – we have implemented this change.

Page 9 - "for spiking a feedforward network" - what does this mean - do they mean "for spiking in a feedforward network"? Aren't these other network mechanisms less biological realistic than the one presented here? I'd like to see some critical comparison between the models.

Thank you for spotting this, this was actually a typo: the sentence should read “for a spiking feedforward network”, which in this case semantically alters the meaning.

Page 9 - "makes a clear prediction...should impact subsequent navigation and the formation of successor features" - This is not a clear prediction but is instead circular - it essentially says - "if successor representations are not formed successor representations will not be observed" This is not much use to an experimentalist. This prediction should be stated in terms of a clear experimental prediction that refers only to physical testable quantities in an experiment and not circularly referring to the same vague and abstract concept of successor representations.

Page 9 "Lesions of the medial septum" - inactivation of the medial septum has also been shown to impair performance in Morris water maze (Chrobak et al. 2006).

We have addressed both of these points with changes to the same paragraph, so we have condensed them for readability. Firstly, we agree our stated “clear prediction” of the model was, in fact, unclear. We have rewritten the paragraph (see below) to clarify what we meant by this. Further, we were unable to locate the *Chrobak et al., 2006* reference, but found a *Chrobak et al., 1989* that matches this description. This is indeed relevant and we have added a citation (let us know if this was not the intended reference or if there is an additional relevant one):

Chrobak, J. J., Stackman, R. W., & Walsh, T. J. (1989). Intraseptal administration of muscimol produces dose-dependent memory impairments in the rat. *Behavioral & Neural Biology, 52*(3), 357–369. https://doi.org/10.1016/S0163-1047(89)90472-X

However, we noted that this paper uses a Muscimol inactivation to medial septum, which was shown by Bolding et al 2019 to disrupt place-related firing as well as theta-band activity, so it is possible that the disruption to place code is what is driving the navigational deficit. Also, we accidentally referred to the inactivations performed by Bolding and colleagues as lesions, but in fact they performed temporary inactivations with a variety of drugs (tetracaine, muscimol, gabazine; the latter of which disrupted theta but left place-related firing intact).

We have modified our paragraph describing these points and the predictions of our model as follows:

“Our theory makes the prediction that theta contributes to learning predictive representations, but is not necessary to maintain them. Thus, inhibiting theta oscillations during exposure to a novel environment should impact the formation of successor features (e.g., asymmetric backwards skew of place fields) and subsequent memory-guided navigation. However, inhibiting theta in a familiar environment in which experience-dependent changes have already occurred

should have little effect on the place fields: that is, some asymmetric backwards skew of place fields should be intact even with theta oscillations disrupted. To our knowledge this has not been directly measured, but there are some experiments that provide hints. Experimental work has shown that power in the theta band increases upon exposure to novel environments [62] – our work suggests this is because theta phase precession is critical for learning and updating predictive maps for spatial navigation. Furthermore, it has been shown that place cell firing can remain broadly intact in familiar environments even with theta oscillations disrupted by temporary inactivation or cooling [63, 64]. It is worth noting, however, that even with intact place fields, these theta disruptions impair the ability of rodents to reach a hidden goal location that had already been learned, suggesting theta oscillations play a role in navigation behaviours even after initial learning [63, 64]. Other work has also shown that muscimol inactivations to medial septum can disrupt acquisition and retrieval of the memory of a hidden goal location [65, 66], although it is worth noting that these papers use muscimol lesions which Bolding and colleagues show also disrupt place-related firing, not just theta precession."

> *Page 9 - "to reach a hidden goal" - A completely different hippocampal modeling framework was used to model the finding of hidden goals in the Morris water maze in Erdem and Hasselmo, 2012, Eur. J. Neurosci and earlier work by Redish and Touretzky 1998, Neural Comp. To clarify the status of the successor representation framework relative to these older models that do not use successor representations, it would be very useful to have a few sentences of discussion about how the successor representation differs and is somehow either advantageous or biologically more realistic than these earlier models.*

We agree this would be helpful, and have added the following text to the discussion:

"A number of other models describe how physiological and anatomical properties of hippocampus may produce circuits capable of goal-directed spatial navigation [30, 27, 23]. These models adopt an approach more characteristic of model- based RL, searching iteratively over possible directions or paths to a goal [30] or replaying sequences to build an optimal transition model from which sampled trajectories converge toward a goal [27] (this model bears some similarities to the SR that are explored by [40], which shows that under certain assumptions, dynamics converge to SR under a similar form of learning). These models rely on dynamics to compute the optimal trajectory, while the SR realises the statistics of these dynamics in the rate code and can therefore adapt very efficiently. Thus, the SR retains some efficiency benefits. The models cited above are very well-grounded in known properties of hippocampal physiology, including theta precession and STDP, whereas until recently, SR models have enjoyed a much looser affiliation with exact biological mechanisms. Thus, a primary goal of this work is to explore how hippocampal physiological properties relate to SR learning as well."

> *Page 9 - "physical barrier to binding" - this is again very confusing as there is no physical barrier in the hippocampus. They should instead say "anatomical segregation"*

Thank you for the suggestion – we have implemented this change as well.

> *Citation 32 - Mommenejad and Howard, 2018 - This is a very important citation and highly relevant to the discussion. However, I think it should just be cited as BioRXiv. It is confusing to call it a preprint.*

Thank you for highlighting this, we have now changed the citation of this and all other cited preprints to their appropriate server e.g. bioRxiv.

Reviewer #2 (Public Review):

The authors present a set of simulations that show how hippocampal theta sequences may be combined with spike time-dependent plasticity to learn a predictive map - the successor representation - in a biologically plausible manner. This study addresses an important question in the field: how might hippocampal theta sequences be combined with STDP to learn predictive maps? The conclusions are interesting and thought-provoking. However, there were a number of issues that made it hard to judge whether the conclusions of the study are justified. These concerns mainly surround the biological plausibility of the model and parameter settings, the lack of any mathematical analysis of the model, and the lack of direct quantitative comparison of the findings to experimental data.

While the model uses broadly realistic biological elements to learn the successor representation, there remain a number of important concerns with regard to the biological plausibility of the model. For example, the model assumes that each CA3 cell connects to exactly 1 CA1 cell throughout the whole learning process so that each CA1 cell simply inherits the activity of a single CA3 cell. Moreover, neurons in the model interact directly via their firing rate, yet produce spikes that are used only for the weight updates. Certain model parameters also appeared to be unrealistic, for example, the model combined very wide place fields with slow running speeds. This leaves open the question as to whether the proposed learning mechanism would function correctly in more realistic parameter settings. Simulations were performed for a fixed running speed, thereby omitting various potentially important effects of running speed on the phase precession and firing rate of place cells. Indeed, the phase precession of CA1 place cells was not shown or discussed, so it is unclear as to whether CA1 cells produce realistic patterns of phase precession in the model.

The fact that a successor-like representation emerges in the model is an interesting result and is likely to be of substantial interest to those working at the intersection between neuroscience and artificial intelligence. However, because no theoretical analysis of the model was performed, it remains unclear why this interesting correspondence emerges. Was it a coincidence? When will it generalise? These questions are best answered by mathematical analysis of the model (or a reduced form of it).

Several aspects of the model are qualitatively consistent with experimental data. For example, CA1 place fields clustered around doorways and were elongated along walls. While these findings are important and provide some support for the model, considerable work is required to draw a firm correspondence between the model and experimental data. Thus, without a quantitative comparison of the place field maps in experimental data and the model, it is hard to draw strong conclusions from these findings.

Overall, this study promises to make an important contribution to the field, and will likely be read with interest by those working in the fields of both neuroscience and artificial intelligence. However, given the above caveats, further work is required to establish the biological plausibility of the model, develop a theoretical understanding of the proposed learning process, and establish a quantitative comparison of the findings to experimental data.

Thank you for the positive comments about the work, and for the detailed and constructive review. We appreciate the time spent evaluating the model and understanding its features at a deep level. Your comments and suggestions have led to exciting new simulation results and a theoretical analysis which shed light on the connections between TD learning, STDP and phase precession.

We have incorporated a number of new simulations to tackle what we believe are your most pressing concerns surrounding the model's biological plausibility. As such, we have extended the hyperparameter sweep (Fig. 2 Supp 3) to include the phase precession parameters you recommended, as well as three new multipanel supplementary figures satisfying your recommendations (Fig 2. Supps 1, 2 & 4). Collectively, these figures show that the specifics of our results, which as you pointed out might have been produced with biologically implausible values (place cell size, movement speed/statistics, weight initialisation, weight updating schedule and phase precession parameters), do not fundamentally depend on the specific values of these parameters: the mechanism still learns predictive maps close in form to the TD successor features. In the hyperparameter sweep, we do find that results are sensitive to specific parameter values (Supp. Fig 3), but that interestingly, the optimal values of these parameters are remarkably close to those observed experimentally. We have also written an extensive new theory section analysing why theta sequences plus STDP approximates TD learning. In addition the methods section has been added to and reordered to make some of the subtler aspects of our model (i.e. the mapping of rates-to-rates and weight fixing during learning) more clear.

At a high level, regarding our claim of biological plausibility, we like to clarify our intended contribution and give context to some responses below. We have added the following paragraph to the discussion in order to accurately represent the scope of our work:

"While our model is biologically plausible in several respects, there remain a number of aspects of the biology that we do not interface with, such as different cell types, interneurons and membrane dynamics. Further, we do not consider anything beyond the most simple model of phase precession, which directly results in theta sweeps in lieu of them developing and synchronising across place cells over time [60]. Rather, our philosophy is to reconsider the most pressing issues with the standard model of predictive map learning in the context of hippocampus (e.g., the absence of dopaminergic error signals in CA1 and the inadequacy of synaptic plasticity timescales). We believe this minimalism is helpful, both for interpreting the results presented here and providing a foundation for further work to examine these biological intricacies, such as the possible effect of phase offsets in CA3, CA1 [61] and across the dorsoventral axis [62, 63], as well as whether the model's theta sweeps can alternately represent future routes [64] e.g. by the inclusion of attractor dynamics [65]."

Reviewer #2 (Recommendations for the authors):

> *This is an interesting study, and I enjoyed reading it. However, I have a number of concerns, particularly regarding the biological plausibility of the model, that I believe can be addressed with additional simulations and analysis.*

Thank you again for your thorough appraisal of our work. Your suggestions have led to new simulations and analyses that have contributed to a significantly improved manuscript. To briefly summarise, these include: 3 new multipanel supplementary figures examining the effects of place field size, running speed, phase precession parameters, weight initialisation,weight update regimes and CA1 phase precession; a new appendix providing theoretical analyses and insight into how and why the model approximates temporal difference learning; and an extension of the hyperparameter sweep analysis to include the parameters controlling phase precession.

> *Major comments:*

- I had a number of concerns regarding the biological plausibility of the model and the choice of parameter settings, especially:

1) Mapping from rates to rates. The CA3 neurons act on CA1 neurons via their firing rate rather than their spikes, but the STDP rule acts on the spikes. What happens if the CA1 neurons are driven by the synaptically-filtered CA3 spikes rather than the underlying rates? How does the model perform, and how does the performance vary with the number of CA3 neurons (since more neurons may be required in order to average over the stochastic spikes)?

We agree that swapping rates for spikes would move the model in the direction of being more biologically plausible; however, this ends up complicating the central comparison of the work. The purpose of this study was to test the hypothesis that a combination of STDP and theta phase precession can approximate the learning of successor representations via temporal difference (TD) learning. As such, since this TD learning rule applies to continuous firing rate values (e.g. de Cothi & Barry 2020), we find this mapping of rates to rates is an essential component to facilitate fair comparison between the two learning rules. This also simplifies our model and its interpretation, as it allows us to avoid the complexity of spiking models. However, we recognise that this is a biologically implausible assumption that we are making. An avenue for correcting this in future work would be to adopt the approach of Brea et al 2016 or Bono et al 2021 (on bioRxiv, also currently in review at eLife). We have now added the following text to the beginning of the results section to clarify why this particular set up was used and its caveats:

"Further, the TD successor matrix M_{ij} can also be used to generate the 'TD successor features' ... allowing for direct comparison and analyses with the STDP successor features (Eqn. 2), using the same underlying firing rates driving the TD learning to sample spikes for the STDP learning. This abstraction of biological detail avoids the challenges and complexities of implementing a fully spiking network, although an avenue for correcting this would be the approach of Brea et al., 2016 and Bono et al., 2021 [41, 43])."

2) Weights are initialised as Wij=deltaij, meaning a 1-1 correspondence from CA3 to CA1 cells. This would have been ok, except that the weights are not updated during learning - they are held fixed during the entire learning phase and only updated on aggregate after learning. Thus, during the entire learning process each CA1 cell is driven by exactly 1 CA3 cell, and therefore simply inherits (or copies) the activity of that CA3 cell (according to equation 2). If either 1) a more realistic weight initialisation were used (e.g., random) or 2) weights were updated online during learning, it seems likely that the proposed mechanism would no longer work.

Thank you for this suggestion. Originally the 1-1 correspondence from CA3 to CA1 cells was to directly correspond to the definition of a successor feature (in which each successor feature corresponds to the predicted activity of a specific basis feature, e.g. Stachenfeld *et al*., 2017; de Cothi & Barry 2020). However we acknowledge the biological implausibility of this approach. As such, we have updated the manuscript to include analyses of simulations where both the target CA1 activity is initialised by random weights (i.e. not the identity matrix), as well as where this target activity is updated online during learning (Fig. 2 Supp 2). As we show, neither manipulation inhibits successful learning of the STDP successor features, with the caveat that when updating the target weights online, the target features need to be partially anchored to the external world to prevent perpetual drift in the target population. We now summarise these new simulations in the results section:

"This effect is robust to variations in running speed (Fig.2–Supplement 1b) and field sizes (Fig. 2–Supplement 1c), as well as scenarios where target CA1 cells have multiple firing fields (Fig. 2–Supplement 2a) that are updated online during learning (Fig. 2–Supplement 2bc; see Supplementary Materials for more details)"

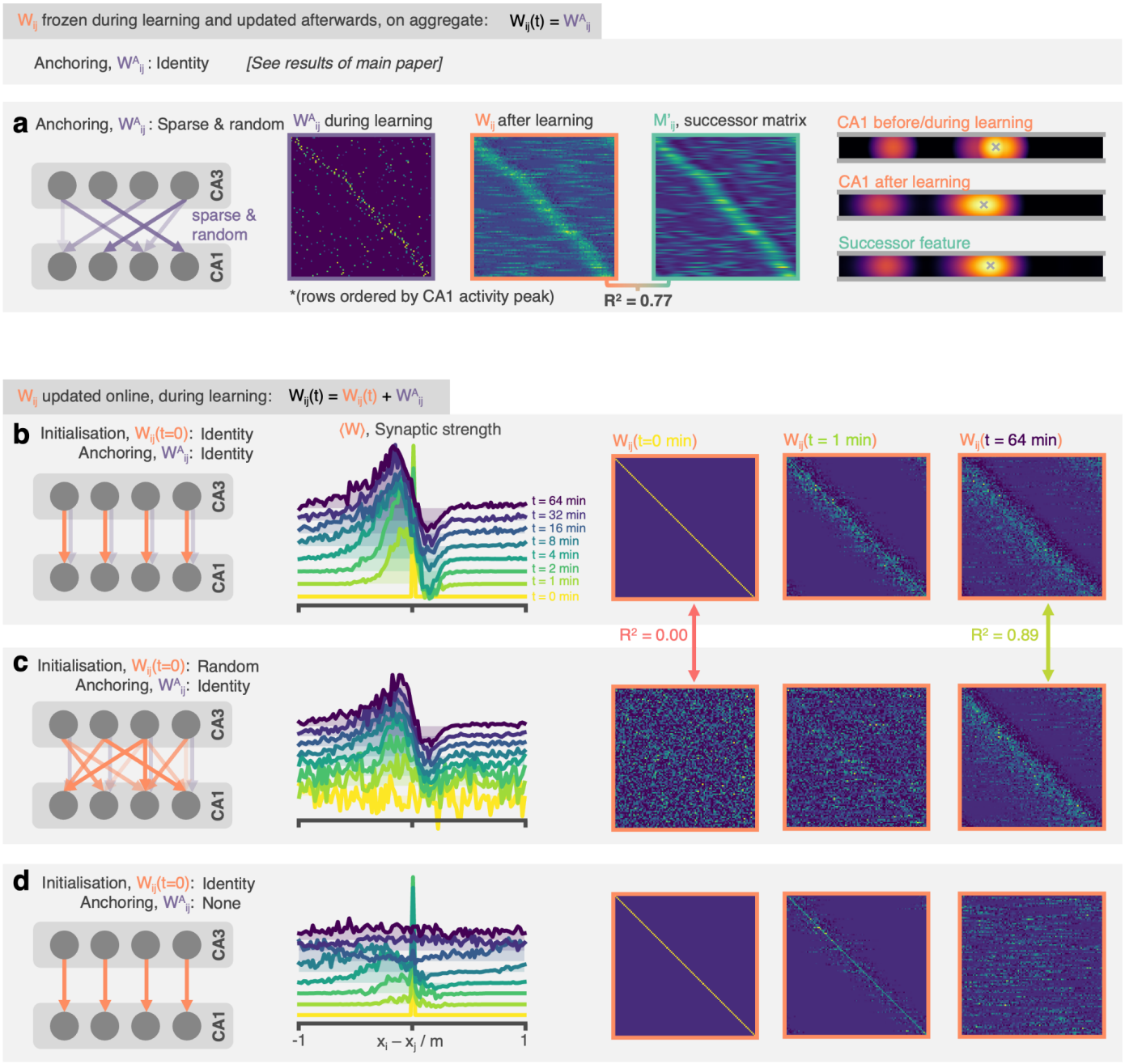


Figure 2 supplementary 2: *The STDP and phase precession model learns predictive maps irrespective of the weight initialisation and the weight updating schedule. In the original model weights are set to the identity before learning and kept ("anchored") there, only updated on aggregate after learning. In these panels we explore variations to this set-up. a (Left) Weights are anchored to a sparse random matrix, not the identity. (Middle) Three weight matrices show the random weights before/during learning, the weights once they have been updated on aggregate after learning and the successor matrix corresponding to the successor features of the mixed features. Matrix rows are ordered by peak CA1 activity location in order that some structure is visible. (Right) An example CA1 feature (top) before learning and (middle) after learning alongside (bottom) the corresponding successor feature. b (Left) The weight matrix is no longer fixed during learning, instead it is initialised to the identity and updated online during learning. A fixed component (0.5 x δij) is added to "anchor" the downstream representations. (Middle and right) After learning the STDP weights show an asymmetric shift and skew against the direction of motion and a negative band ahead of the diagonal just as was observed for successor matrices and the fixed weight model. This backwards expansion does not carry on extending indefinitely (a risk when the weights are updated online) but stabilises. c Like panel b but weights are randomly initialised. After learning the weights have "forgotten" their initial structure and are essentially identical to in the case of identity initialisation. d Like panel b except no anchoring weights are added. Now there is no fixed component anchoring CA1 representations, structure in the synaptic weights rapidly disintegrates.*

and elaborate on this method in the appendices/methods:

*"**Random initialisation**: In figure 2 supplement 2, panel a, we explore what happens if weights are initialised randomly. Rather than the identity, the weight matrix during learning is fixed ("anchored") to a sparse random matrix W_A; this is defined such that each CA1 neuron receives positive connections from 3, 4 or 5 randomly chosen CA3 neurons with weights summing to one. In all other respects learning remains unchanged. CA1 neurons now have multi-modal receptive fields since they receive connections from multiple, potentially far apart, CA3 cells. This shouldn't cause a problem since each sub-field now acts as its own place field phase precessing according to whichever place cells in CA3 is driving it. Indeed it doesn't: after learning with this fixed but random CA3-CA1 drive, the synaptic weights are updated on aggregate and compare favourably to the successor matrix (panel a, middle and right). Specifically this is the successor matrix which maps the unmixed uni-modal place cells in CA3 to the successor features of the new multi-modal "mixed" features found in CA1 before learning. We note in passing that this is easy to calculate due to the linearity of the successor feature (SF): a SF of a linear sum of features is equal to a linear sum of SF, therefore we can calculate the new successor matrix using the same algorithm as before (described in the methods) then rotating it by the sparse random matrix* $\mathsf{M}'_{ij} = \sum_k \mathsf{W}^{\mathsf{A}}_{ik} \mathsf{M}_{kj}$.

In order that some structure is visible matrix rows (which index the CA1 postsynaptic cells) have been ordered according to the location of the CA1 peak activity. This explains why the random sparse matrix (panel a, middle) looks ordered even though it isn't. After learning the STDP successor feature looks close in form to the TD successor feature and both show a shift and skew backwards along the track (panel a, rights, one example CA1 field shown)."

***Online weight updating:** In Fig. 2 supplement 2, panels b, c and d, we explore what happens if the weights are updated online during learning. It is not possible to build a stable fully online model (as we suspect the review realised) and it is easy to understand why: if the weight matrix doing the learning is also the matrix doing the driving of the downstream features then there is nothing to prevent instabilities where, for example, the downstream feature keeps shifting backwards (no convergence) or the weight matrix for some/all features disappears or blows up (incorrect convergence). However it is possible to get most of the way there by splitting the driving weights into two components. The first and most significant component is the STDP weight matrix being learned online, this creates a "closed loop" where changes to the weights affects the downstream features which in turn affect learning on the weights. The second smaller component is what we call the "anchoring" weights, which we set to a fraction of the identity matrix (here 1/2) and are not learned. In summary, Eqn. (16) becomes*

$$\tilde{\psi}_i(\mathbf{x}, t) = \sum_j \left(\mathsf{W}_{ij}(t) + \mathsf{W}^{\mathsf{A}}_{ij} \right) f_j(\mathbf{x}, t)$$

These anchoring weights provide structure, analogous to a target signal or "scaffold" onto which the successor features will learn without risk of infinite backwards expansion or weight decay. After learning when analysing the weight/successor features the anchoring component is not considered.

This is not a hack: every other model of TD learning implicitly or explicitly has a form of anchoring. For example in classical TD learning each successor feature receives a fixed "reward" signal from the feature it is learning to predict (this is the second term in equation (23) of our methods). Even other "synaptically plausible" models include a non-learnable constant drive (see Bono et al.'s[41] CA3-CA1 model, more specifically the bias term in their Eqn. (12)). This is the approach we take here. We add the additional constraint that the sum of each row of the weight matrix must be smaller than or equal to 1, enforced by renormalisation on each time step. This constraint encodes the notion that there may be an energetic cost to large synaptic weight matrices and prevents infinite growth of the weight matrix

$$\mathsf{W}_{ij}(t) \leftarrow \frac{\mathsf{W}_{ij}(t)}{\max(1, \sum_j \mathsf{W}_{ij})}$$

The resulting evolution of the learnable weight component, $W_{ij}(t)$, is shown in panel b (middle shows row aligned averages of $W_{ij}(t)$ from t=0 minutes to to = 64 minutes, on the full matrices are shown) and panel f (full matrix) from being initialised to the identity. The weight matrix evolves to look like a successor matrix (long skew left of diagonal, negative right of diagonal). One risk, when weights are updated online, is that the asymmetric expansion continues indefinitely. This doesn't happen and the matrix stabilises after 15 minutes (panel e, color progression). It is important

to note that the anchoring component is smaller than the online weight component and we believe it could be made very small in the limit of less noisy learning (e.g. more cells or higher firing rates).

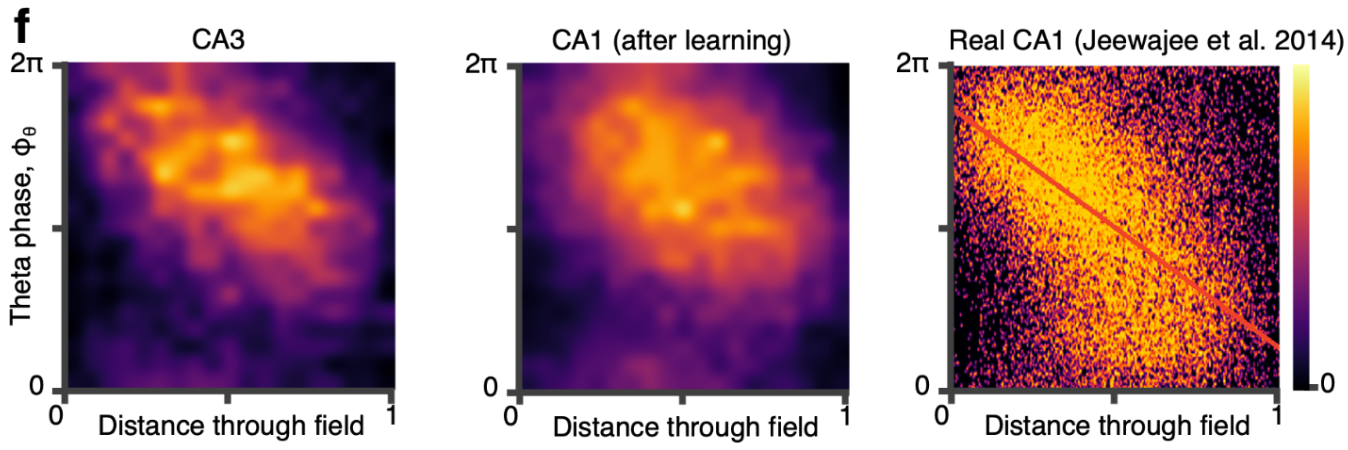
In panel c we explore the combination: random weight initialisation and online weight updating. As can be seen, even with rather strong random initial weights learning eventually "forgets" these and settles to the same successor matrix form as when identity initialisation was used.

In panel d we show that anchoring is essential. Without it ($W_{Aij} = 0$) the weight matrix initially shows some structure shifting and skewing to the left but this quickly disintegrates and no observable structure remains at the end of learning.

One interpretation of our set-up (the original one, described in the main text of the paper where weights are not updated online) is that it matches the "Separate Phases of Encoding and Retrieval Model" model [Hasselmo (2002)]. This paper describes how LTP between CA1 and CA3 synapses is strongest at the phase of theta when input to CA1 is primarily coming from entorhinal cortex. To quote the abstract of this paper: *"effective encoding of new associations occurs in the phase when synaptic input from entorhinal cortex is strong and long-term potentiation (LTP) of excitatory connections arising from hippocampal region CA3 is strong, but synaptic currents arising from region CA3 input are weak".* Broadly speaking, this matches what we have here. That is to say: what drives CA1 during learning are not the synapses onto which learning is accumulating. Of course we don't replicate this model in all its details – for example we don't actually separate CA1 drive into two phases, and don't model phase dependent LTD and so don't reproduce their memory extinction results – but, philosophically, it is similar.

> *3) Lack of discussion of phase precession in CA1 cells. What are the theta firing patterns of CA1 (successor) cells in the model? Do they exhibit theta sequences and/or phase precession? We are never told this. The spike phase of the downstream CA1 cell is extremely important for STDP, as it determines whether synapses associated with past or future events are potentiated or suppressed (see Figure 8 of Chadwick et al. 2016, eLife). Based on my understanding, in the current setup CA1 place cells should produce phase precession during learning (before weights are updated), but only because each CA1 cell copies the activity of exactly one CA3 cell, which is unrealistic. Moreover, after the weights are updated, whether they produce phase precession is no longer clear. It is important to determine whether the proposed mechanism works in the more realistic scenario in which both CA3 and CA1 cells exhibit phase precession, but CA1 cells are driven by multiple CA3 cells.*

Thank you for these suggestions. We now show in Fig. 2 supplement 4f that the CA1 STDP successor features in the model do indeed inherit this phase precession:



CA1 cells will phase precess when driven by multiple CA3 place cells. Here we show phase precession (spike probability for different theta phases against distance travelled through field) for CA3 basis features and CA1 STDP successor features after learning. Although noisier there is still a clear tendency for CA1 cells to phase precess. Real

CA1 cell phase precession can be 'noisy'; we show for comparison a phase precession plot for CA1 place field taken from Jeewajee et al. (2014), the same data for which we fitted our parameters.

The reason for this is that CA1 cells are still localised and therefore driven mostly by cells in CA3 which are close and which peak in activity together at a similar phase each theta cycle. As the agent moves through the CA1 cell it also moves through all the CA3 cells and their peak firing phase 'precesses' driving an earlier peak in the CA1 firing. Phase precession is CA1 after learning is noisier/broader than CA3 but far from non-existent and looks similar to real phase precession data from cells in CA1. This result is described in the main text:

"In particular, the parameters controlling phase precession in the CA3 basis features (Fig. 2–supplement 4a) can affect the CA1 STDP successor features learnt, with 'weak' phase precession resembling learning in the absence of theta modulation (Fig. 2–supplement 4bc), biologically plausible values providing the best match to the TD successor features (Fig. 2–supplement 4d) and 'exaggerated' phase precession actually hindering learning (Fig. 2–supplement 4e; see Supplementary Materials for more details). Additionally, we find these CA1 cells go on to inherit phase precession from the CA3 population even after learning when they are driven by multiple CA3 fields (Fig. 2–supplement 4f)."

And we elaborate on this in the appendices/methods:

"***Phase precession of CA1:*** *In most results shown in this paper the weights are anchored to the identity during learning. This means each CA1 cells inherits phase precession from the one and only one CA3 cell it is driven by. It is important to establish whether CA1 still shows phase precession after learning when driven by multiple CA3 cells or, equivalently, during learning when the weights aren't anchored and it is therefore driven by multiple CA3 neurons. Analysing the spiking data from CA1 cells after learning (phase precession turned on) shows it does phase precession. This phase precession is noisier than the phase precession of a cell in CA3 but only slightly and compares favourably to real phase precession data for CA1 neurons (panel f, right, with permission from Jeewajee et al. (2014) [46]).*

The reason for this is that CA1 cells are still localised and therefore driven mostly by cells in CA3 which are close and which peak in activity together at a similar phase each theta cycle. As the agent moves through the CA1 cell it also moves through all the CA3 cells and their peak firing phase precesses driving an earlier peak in the CA1 firing. Phase precession is CA1 after learning is noisier/broader than CA3 but far from non-existent and looks similar to real phase precession data from cells in CA1."

Additionally, by extending our parameter sweep to include phase precession parameters (Fig. 2–supplement 3 panel c, last 2 subplots), we now show that the biologically derived values for the parameters determining the phase precession in the model are in fact optimally placed to approximate the TD learning of successor features (Fig. 2–supplement 4, please see response to point 5 for more details).

Finally, we show that the CA1 successor features can still be successfully learnt via the STDP + phase precession mechanism when the target features are driven by multiple CA3 cells (Fig. 2 supplement 2A), and when the target features are updated by the learnt weights online (Fig. 2 supplement 2bc, please see response to point 2 for technical details).

> *4) Related to the preceding comment, there is a phase shift/delay between CA3 and CA1 (Mizuseki, Buzsaki et al., 2010). This doesn't seem to have been taken into account. Can the model be set up so that i) CA1 cells receive inputs from multiple CA3 cells ii) both CA3 and CA1 cells exhibit phase precession iii) there is the appropriate phase delay between CA3 and CA1?*

Thank you for this comment, as it provoked much thought. At the level of individual cells in our model, the phase shift presented by Mizuseki, Buzsaki et al., 2010 (i.e. CA1 being shifted temporally just ahead of CA3) is functionally near-identical to if each CA3 basis feature were connected to a different CA1 cell

slightly further ahead of it down the track. Therefore, in total, this would simply manifest as a rotation on the weight matrix (e.g. realignment of CA1 cells along the track). Thus perhaps these phase delays are important for other aspects of learning we are not capturing here. However, if this shift were more substantial, it is not entirely clear what would happen. We identify this as a limitation and direction for future work in the new paragraph we have added that discussing the limits of the model's biological plausibility (reprinted below for convenience):

"While our model is biologically plausible in several respects, there remain a number of aspects of the biology that we do not interface with, such as different cell types, interneurons and membrane dynamics. Further, we do not consider anything beyond the most simple model of phase precession, which directly results in theta sweeps in lieu of them developing and synchronising across place cells over time [60]. Rather, our philosophy is to reconsider the most pressing issues with the standard model of predictive map learning in the context of hippocampus (e.g., the absence of dopaminergic error signals in CA1 and the inadequacy of synaptic plasticity timescales). We believe this minimalism is helpful, both for interpreting the results presented here and providing a foundation for further work to examine these biological intricacies, such as the possible effect of phase offsets in CA3, CA1 [61] and across the dorsoventral axis [62, 63], as well as whether the model's theta sweeps can alternately represent future routes [64] by the inclusion of attractor dynamics [65]."

> *5) Dependence of learning on the noisiness of phase precession. The hyperparameter sweep seems to omit some of the most important variables, such as the spread parameter (kappa) and the place field width and running speed (see next comment). Since the successor representation is shown to be learned well when kappa=1 but not when kappa=0 (i.e. when phase precession is removed), this leaves open the question of what happens when kappa is bigger than or small than 1. It would be nice to see kappa systematically varied and the consequences explored.*

Thank you for this suggestion. We have now extended our parameter sweep (Fig. 2 supplement 3) to systematically determine the effect of variations in the noisiness of the phase precession (kappa) and the proportion of the theta cycle in which the precession takes place (beta). Interestingly, we find that the biologically derived parameters are in fact optimally placed to approximate the TD learning of successor features (Fig. 2 supplement 3c & 4a-e). We summarise these results in the main text:

"In particular, the parameters controlling phase precession in the CA3 basis features (Fig. 2–supplement 4a) can affect the CA1 STDP successor features learnt, with 'weak' phase precession resembling learning in the absence of theta modulation (Fig. 2–supplement 4bc), biologically plausible values providing the best match to the TD successor features (Fig. 2–supplement 4d) and 'exaggerated' phase precession actually hindering learning (Fig. 2–supplement 4e; see Supplementary Materials for more details). Additionally, we find these CA1 cells go on to inherit phase precession from the CA3 population (Fig. 2–supplement 4f)."

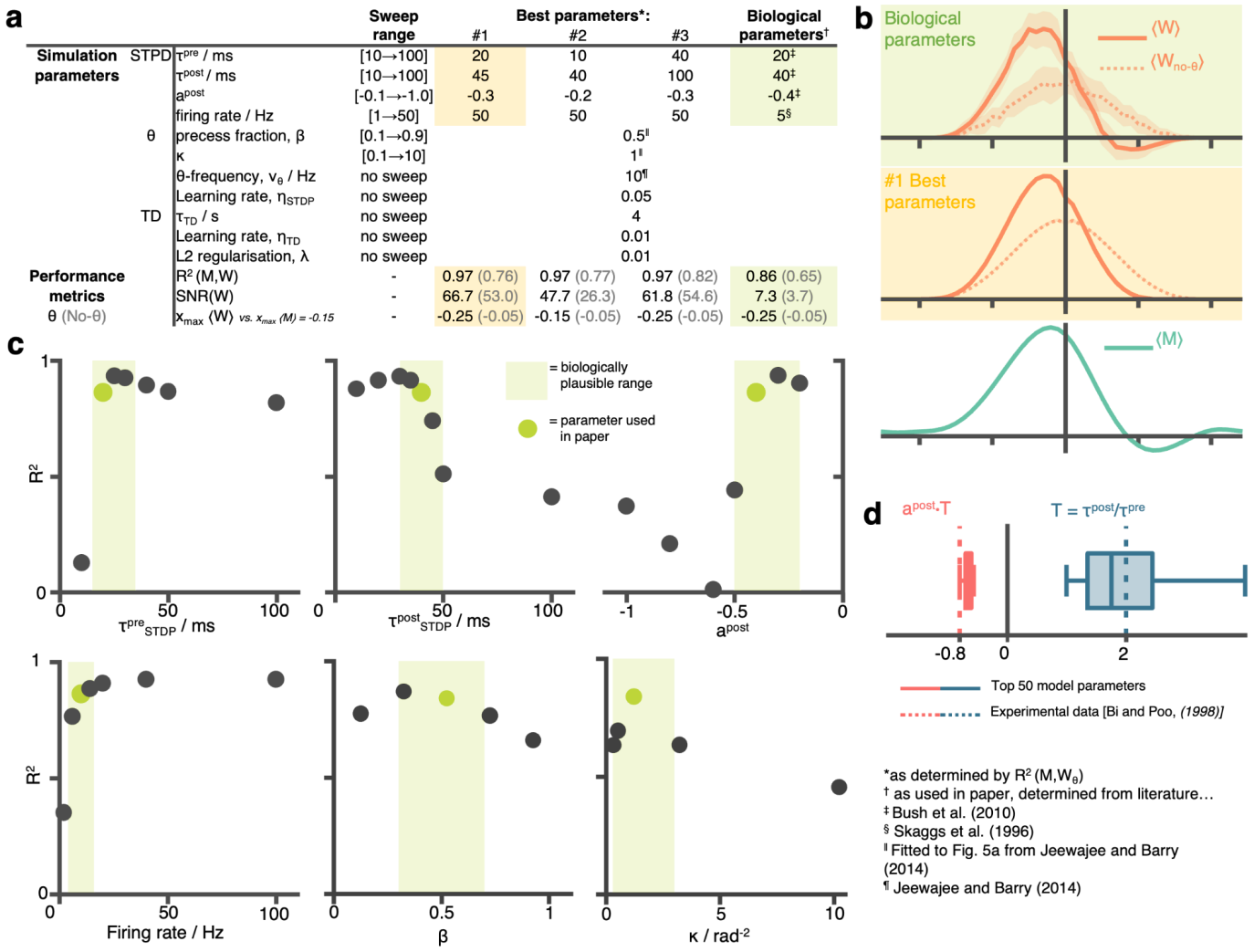
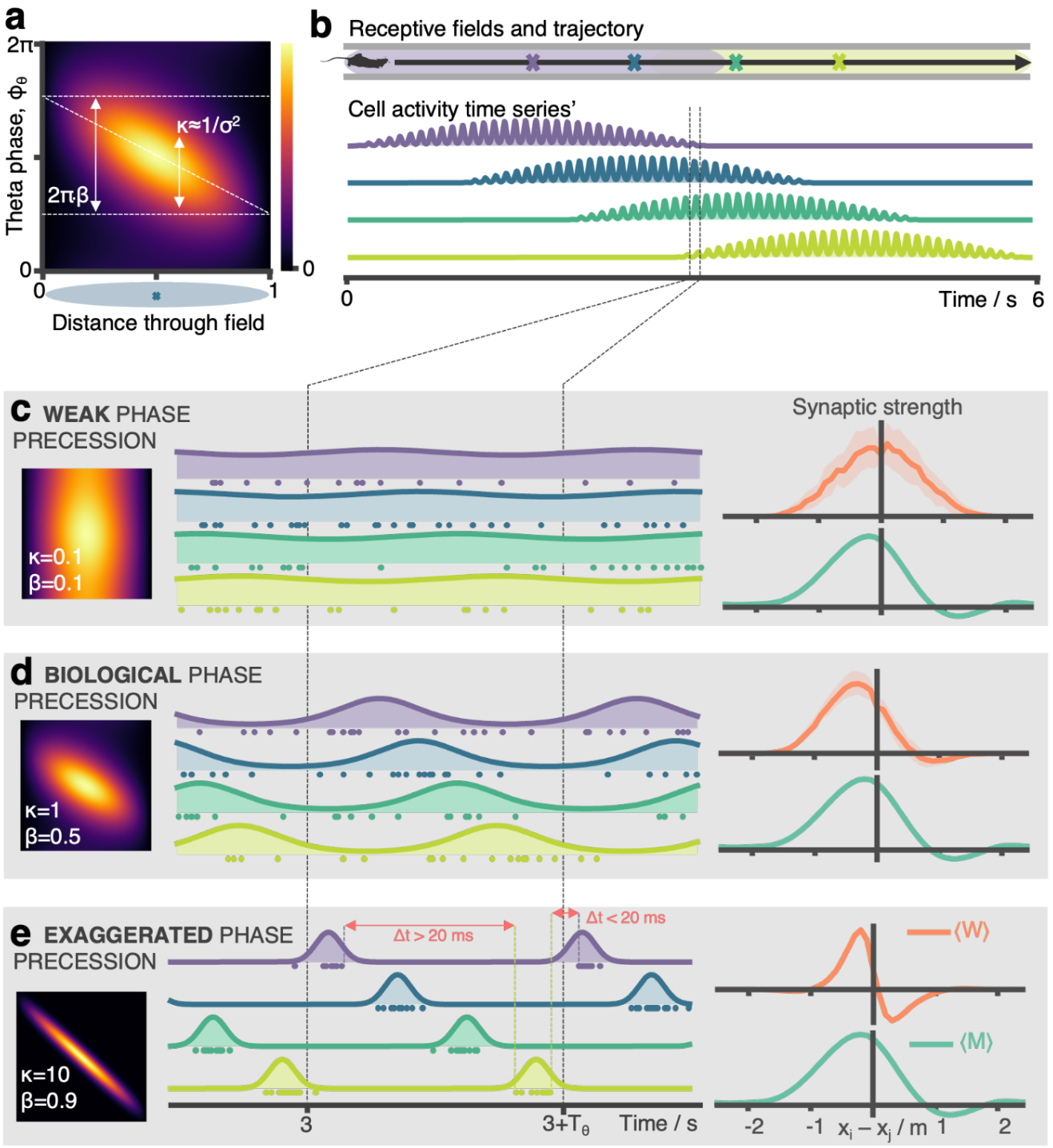


Figure 2 supplement 3: *A hyperparameter sweep over STDP and phase precession parameters shows that biological parameters are suffice, and are near-optimal for approximating the successor features a A table showing all parameters used in this paper and the ranges over which the hyperparameter sweep was performed. For each parameter setting we estimate performance metrics to judge whether the STDP parameters do well at learning the successor features. b Visually inspecting the row aligned STDP weight matrices we see the optimal parameters do not significantly out perform the biologically chosen ones. Although the optimal parameter setting results in a slightly higher R2, they fail to capture the right-of-centre negative weights present in the TD successor matrix, unlike the biological ones. c Slices through the parameter sweep hypercube. For each plot, parameter values of the other five variables are fixed to the green values (i.e. are the ones used in this paper). d The top 50 performing parameter combination are stored and box plots for the conjugate parameter $T = \tau_{pre}$, τ_{post} the ratio of time windows for potentiation and depression, and $-a_{post} \cdot T$, effectively the ratio of the areas under the curve left and right of the y-axis on the STDP plot Fig. 1b. In both cases the 'best parameters' include the true parameter values, measured experimentally by Bi and Poo (1998) [19].*

In an additional supplementary figure (Fig. 2–supplement 4) we delve into these hyperparameter sweep results showing examples of too-much or too-little phase precession on the learnt successor features and attempt to shed light on why this intermediate optima exist.



Supplementary Figure 4: *Biological phase precession parameters are optimal for learning the SR. a We model phase precession as a von Mises centred at a preferred theta phase which precesses in time. This factor modulates the spatial firing field. It is parameterised by κ (von Mises width parameter, aka noise) and β (fraction of full 2π phase being swept, diagonal line). We showed in a previous figure that biological phase precession parameters are optimal. Any more or less phase precession degrades performance. It is easy to understand why: b Consider four place cells on a track (purple, blue, green, yellow) where the first and last just overlap. c In the weak phase precession regime there is no ordering to the spikes and STDP can't learn the asymmetry in the successor matrix (right) d In the medium phase precession regime spikes are broadly ordered in time (purple then blue then green...) so the symmetry is broken and STDP learns a close approximation the successor matrix e) In the "exaggerated" phase precession regime there exist two problems for learning SRs: "causal" bindings (e.g. from presynaptic purple to postsynaptic yellow, which sits in front of purple) are inhibited for anything except the most closely situated cell pairs due to the sharp tuning curves. Secondly, though this is a less important effect, when β is too large it is possible for incorrect "acasual" bindings to be formed due to one cell (e.g. yellow) firing late in theta cycle N just before another cell located far behind it on the track fires (e.g. purple) in theta cycle N+1. f CA1 cells will phase precess when driven by multiple CA3 place*

cells. Here we show phase precession (spike probability for different theta phases against distance travelled through field) for CA3 basis features and CA1 STDP successor features after learning. Although noisier there is still a clear tendency for CA1 cells to phase precess. Real CA1 cell phase precession can be 'noisy'; we show for comparison a phase precession plot for CA1 place field taken from Jeewajee et al. (2014), the same data for which we fitted our parameters.

We also go into further detail in the appendices/methods:

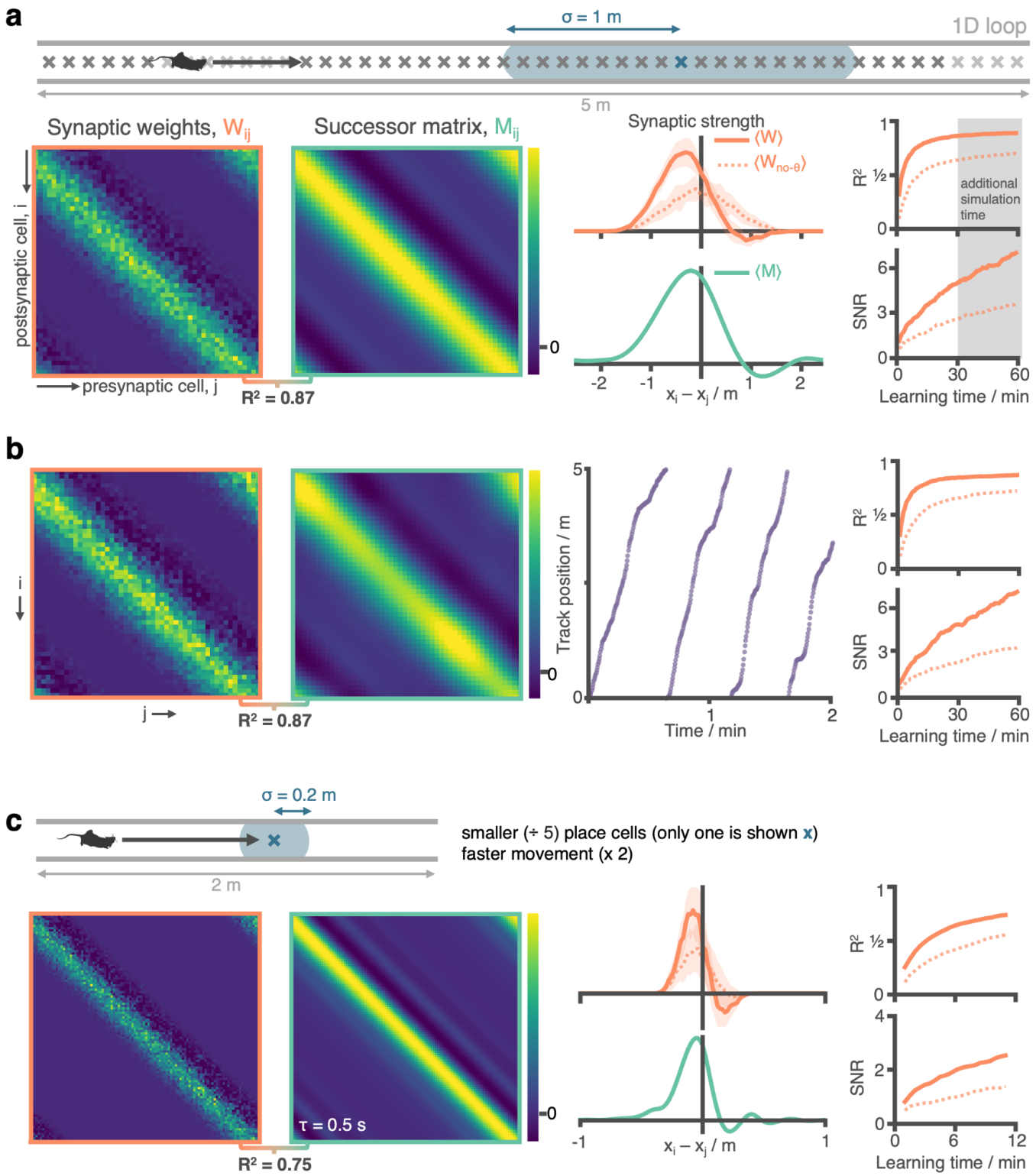
"The optimality of biological phase precession parameters In figure 2 supplement 3 we ran a hyperparameter sweep over the two parameters associated with phase precession: κ, the von Mises parameter describing how noisy phase precession is and β, the fraction of the full 2π theta cycle phase precession crosses. The results show that for both of these parameters there is a clear "goldilocks" zone around the biologically fitted parameters we chose originally. When there is too much (large κ, large β) or too little (small κ, small β) phase precession performance is worse than at intermediate biological amounts of phase precession. Whilst – according to the central hypothesis of the paper – it makes sense that weak or non-existence phase precession hinders learning, it is initially counter intuitive that strong phase precession also hinders learning.

We speculate the reason is as follows, when β is too big phase precession spans the full range from 0 to 2π, this means it is possible for a cell firing very late in its receptive field to fire just before a cell a long distance behind it on the track firing very early in the cycle because 2π comes just before 0 on the unit circle. When κ is too big, phase precession is too clean and cells firing at opposite ends of the theta cycle will never be able to bind since their spikes will never fall within a 20 ms window of each other. We illustrate these ideas in figure 2 supplement 4 by first describing the phase precession model (panel a) then simulating spikes from 4 overlapping place cells (panel b) when phase precession is weak (panel c), intermediate/biological (panel d) and strong (panel e). We confirm these intuitions about why there exists a phase precession "goldilocks" zone by showing the weight matrix compared to the successor matrix (right hand side of panels c, d and e). Only in the intermediate case is there good similarity."

> *6) Wide place fields and slow speeds. Place fields in the model have a diameter of 2 metres. This is quite big - bigger than typical place field sizes in the dorsal hippocampus (which often have around 30 cm diameter, or 15 cm radius). Moreover, the chosen velocity of 16 cm/s is quite slow, and rats often run much faster in experiments (30 cm/s and higher). With the chosen parameters, it takes the rodent 12.5 s to traverse a place field, which is unrealistically long. My concern is that this setup leads to a large number of spikes per pass through a place field and that this unrealistic setting is needed for the proposed mechanism to learn effectively in a reasonable number of laps. What happens when place fields are smaller and running speeds faster, as is typically found in experiments? How many laps are required for convergence?*

Thank you for this suggestion, we now explore this in a new fsupplementary figure, (Fig. 2–supplement 1bc). In summary, we find there is no critical effect on learning with smaller place fields and faster speeds. As hypothesised by the reviewer, we find that the learning is slower (when measured in number of laps) due to the decreased number of spikes, but not with catastrophic effects. This is summarised in the results:

"Thus, the ability to approximate TD learning appears specific to the combination of STDP and phase precession. Indeed, there are deep theoretical connections linking the two - see Methods section 5.8 for a theoretical investigation into the connections between TD learning and STDP learning augmented with phase precession. This effect is robust to variations in running speed (Fig. 2–supplement 1b) and field sizes (Fig. 2–supplement 1c), as well as scenarios where target CA1 cells have multiple firing fields (Fig. 2–supplement 2a) that are updated online during learning (Fig. 2–supplement 2bc; see Supplementary Materials for more details)"



Supplementary Figure 1: *STDP and phase precession combine to make a good approximation of the SR independent of place cell size and running speed statistics. a Figure 2 panels a-e have been repeated (additional 30 minutes simulation carried out) for ease of comparison. b We repeat the experiment with non-uniform running speed. Here, running seed is sampled according to a continuous stochastic process (Ornstein Uhlenbeck) with mean of 16 cm s^{-1} and standard deviation 16 cm s^{-1} thresholded to prevent negative speeds. As can be seen in the trajectory figure speed varies smoothly but significantly, including regions where the agent is almost stationary. Despite this there is no observable difference to the synaptic weights after learning. c We reduce the place cell diameter from 2 m to 0.4 m (5x decrease) and increase the motion speed from 16 cm s^{-1} to 32 cm s^{-1} (2x increase). We increase the cell density along the track from 10 cells m^{-1} to 50 cells m^{-1} to preserve cell overlap density. To reduce the computational load of training we shrink the track length from 5 m to 2 m (any additional track is symmetric and redundant when place cells are this small anyway). Note the adjusted training time: 12 minutes on a 2 m track at 32 cm s^{-1} corresponds to the same number of laps as 60 min on a 5 m track at 16 cm s^{-1} as shown for comparison in panel (a).*

Under these conditions the STDP + phase precession learning rule well approximates the successor features with a shorter time horizon of τ = 0.5.

And elaborated on in the appendices/methods:

*"**Smaller place cells and faster movement:** Nothing fundamental prevents learning from working in the case of smaller place fields or faster movement speeds. We explore this in figure 2 supplement 1, panel c, as follows: the agent speed is doubled from 16 cm s^{-1} to 32 cm s^{-1} and the place field size is shrunk by a factor of 5 from 2 m diameter to 40 cm diameter. To facilitate learning we also increase the cell density along the track from 10 cells m^{-1} to 50 cells m^{-1}. We also shrink the track size from 5 m to 2 m (any additional track is redundant due to the circular symmetry of the set-up and small size of the place cells). We then train for 12 minutes. This time was chosen since 12 minutes moving at 32 cm s^{-1} on a 2 m track means the same number of laps as 60 mins moving at 16 cm s^{-1} on a 5 m track (96 laps in total). Despite these changes the weight matrix converged with high similarity to the successor matrix with a shorter time horizon (0.5 s). Convergence time measured in minutes was faster than in the original case but this is mostly due to the shortened track length and increased speed. Measured in laps it now takes longer to converge due to the decreased number of spikes (smaller place fields and faster movement through the place fields). This can be seen in the shallower convergence curve, panel c (right) relative to panel a."*

> *7) Running speed-dependence of phase precession and firing rate. The rat is assumed to run at a fixed speed - what happens when speed is allowed to vary? Running speed has profound effects on the firing of place cells, including i) a change in their rate of phase precession ii) a change in their firing rate (Huxter et al., 2003). More simulations are needed in which running speed varies lap-by-lap, and/or within laps.*

Thank you for this suggestion, we now explore this in a new supplementary figure, (Fig. 2–supplement 1b, see comment above) where the speed of the rat / agent is allowed to vary smoothly and stochastically. In summary, we find no observable effect on the STDP weight matrix or the TD successor matrix after learning, with the R^2 value between the two. This is summarised in the results:

"Thus, the ability to approximate TD learning appears specific to the combination of STDP and phase precession. Indeed, there are deep theoretical connections linking the two - see Methods section 5.8 for a theoretical investigation into the connections between TD learning and STDP learning augmented with phase precession. This effect is robust to variations in running speed (Fig. 2–supplement 1b) and field sizes (Fig. 2–supplement 1c), as well as scenarios where target CA1 cells have multiple firing fields (Fig. 2–supplement 2a) that are updated online during learning (Fig. 2–supplement 2bc; see Supplementary Materials for more details)"

With further details in the appendices/methods:

*"**Movement speed variability**: Panel b shows an experiment where we reran the simulation shown in paper figures 2a-e except, instead of a constant motion speed, the agent moves with a variable speed drawn from a continuous stochastic process (an Ornstein-Uhlenbeck process). The parameters of the process were selected so the mean velocity remained the same (16 cm s^{-1} left-to-right) but now with significant variability (standard deviation of 16 cm s^{-1} thresholded so the speed can't go negative). Essentially, the velocity takes a constrained random walk. This detail is important: the velocity is not drawn randomly on each time step since these changes would rapidly average out with small dt, rather the change in the velocity (the acceleration) is random - this drives slow stochasticity in the velocity where there are extended periods of fast motion and extended periods of slow motion. After learning there is no substantial difference in the learned weight matrices. This is because both TD and STDP learning rules are able to average-over the stochasticity in the velocity and converge on representations representative of the mean statistics of the motion.*

> *8) Two-dimensional phase precession. There is debate over how 2D environments are encoded in the theta phase (Chadwick et al. 2015, 2016; Huxter et al., 2008; Climer et al., 2013; Jeewajee et al., 2013). This should be mentioned and discussed - how much do the results depend on the*

specific assumptions regarding phase precession in 2D? For example, Huxter et al. found that, when animals pass through the edge of a place field, the cell initially precesses but then processes back to its initial phase, but this isn't captured by the model used in the present study. Chadwick et al. (2016) proposed a model of two-dimensional phase precession based on the phase locking of an oscillator, which reproduces the findings of Huxter et al. and makes different predictions for phase precession in two dimensions than the Jeewajee model used by the authors. It would be nice to test alternative models for 2D phase precession and determine how well they perform in terms of generating successor-like representations.

Thank you for this suggestion. We agree this is an important topic in terms of understanding the correlates and consequences of phase precession. There is a wealth of literature surrounding this topic, some of which we relied upon for defining the model of 2D phase precession implemented here (e.g. Jeewajee et al., 2013 and Chadwick et al. 2015). However, we believe that this would be better suited as a followup to the current study, which addresses the first question of what how closely the representations learned with classical theta precession resemble TD-trained SRs. Rather, we agree that considering alternative 2D models of phase precession would be a wonderful direction for future work and our code is publicly available should anyone wish to explore this.

9) Modelling the distribution of place field sizes along the dorsoventral axis. Two important phenomena were omitted that are likely important and could alter the conclusions. First, there is a phase gradient along the dorsoventral axis, which generates travelling theta waves (Patel, Buszaki et al., 2012; Lebunov and Siapas, 2009). How do the results change when including a 180 (or 360) phase gradient along the DV axis? The authors state that "A consequence of theta phase precession is that the cell with the smaller field will phase precess faster through the theta cycle than the other cell - initially it will fire later in the theta cycle than the cell with a larger field, but as the animal moves towards the end of the small basis field it will fire earlier" - this neglects to consider the phase gradient along the DV axis (see also Leibold and Monsalve-Mecado, 2017). Second, the authors chose three discrete place field sizes for their dorsoventral simulations. How would these simulations look if a continuum of sizes were used reflecting the gradient along the dorsoventral axis? Going further, CA1 cells likely receive input from CA3 cells with a distribution of place field sizes rather than a single place field size - how would the model behave in that case?

Thank you for this interesting point. The model and results presented here pertain more to the role of theta compression (and STDP) in approximating TD learning. However we have now added the following to our discussion to consider these additional aspects of theta oscillations:

“The distribution of place cell receptive field size in hippocampus is not homogeneous. Instead, place field size grows smoothly along the longitudinal axis (from very small in dorsal regions to very large in ventral regions). Why this is the case is not clear – our model contributes by showing that, without this ordering, large and small place cells would all bind via STDP, essentially overwriting the short timescale successor representations learnt by small place cells with long timescale successor representations. Topographically organising place cells by size anatomically segregates place cells with fields of different sizes, preserving the multiscale successor representations. The functional separation of these spatial scales could be further enhanced by a gradient of phase offsets along the dorso-ventral axis, resulting from the theta oscillation being a travelling wave [62, 63]. This may act as a temporal segregation preventing learning between cells of different field sizes, on top of the anatomical segregation we explore here. The premise that such separation is needed to learn multiscale successor representations is compatible with other theoretical accounts for this ordering. Specifically Momennejad and Howard [39] showed that exploiting multiscale successor representations downstream, in order to recover information which is ‘lost’ in the process of compiling state transitions into a single successor representation, typically requires calculating the derivative of the successor representation with respect to the discount parameter. This derivative calculation is significantly easier if the cells – and therefore the successor representations – are ordered smoothly along the hippocampal axis.”

As well as this, we include a new paragraph in the discussion pertaining to these limits in the model's biological plausibility and our intended contribution:

"While the model is biologically plausible in several respects, there remain a number of aspects of the biology that we do not interface with, such as different cell types, interneurons and membrane dynamics. Further, only the most simple model of phase precession is considered, which directly results in theta sweeps in lieu of them developing and synchronising across place cells over time [60]. Rather, our philosophy is to reconsider the most pressing issues with the standard model of predictive map learning in the context of hippocampus. These include the absence of dopaminergic error signals in CA1 and the inadequacy of synaptic plasticity timescales. We believe this minimalism is helpful, both for interpreting the results presented here and providing a foundation on which further work may examine these biological intricacies, such as the possible effect of phase offsets in CA3, CA1 [61] and across the dorsoventral axis [62, 63], as well as whether the model's theta sweeps can alternately represent future routes [64] e.g. by the inclusion of attractor dynamics [65]."

> *- There is no theoretical analysis of why theta sequences+STDP approximates the TD algorithm, or when the proposed mechanism might/might not work. The model is simple enough that some analysis should be possible. It would be nice to see this elaborated on - can a reduced model be obtained that captures the learning algorithm embodied by theta sequences+STDP, and does this reduced model reveal an explicit link to the TD algorithm? If not, then why does it work, and when might it generalise/not work?*

Thank you for this suggestion. We have now updated the manuscript to include a section (Methods 5.8) explaining the theoretical connection between STDP and TD learning. In short, it starts by showing how temporal difference learning can be mathematically recast into a temporally asymmetric Hebbian learning rule reminiscent of simplified STDP. However, in order to recast TD learning in its STDP-like form it is necessary to fix the temporal discount time horizon to the synaptic plasticity timescale. This alone would produce TD-style learning on a time-scale too short to capture meaningful predictions of behaviour. Thus, we show mathematically that the importance of theta phase precession is to provide a precise temporal compression on the input sequences that effectively increases this predictive time horizon from the timescale of synaptic plasticity to the timescale of behaviour. This temporal compression overcomes the timescales problem since, by symmetry, learning a successor feature with a very small time horizon where the input trajectory is temporally compressed is equivalent to learning a successor feature with a long time horizon where the inputs are not compressed. We derive a formula for the amount of compression as a function of the typical speed of a `theta sweep' and estimate a ballpark figure showing that in many cases this compression is enough to extend the synaptic plasticity timescale into behaviourally relevant timescales. In essence, this section provides the mathematics behind the very intuition on which we based the study (e.g. Fig 1). That is:

1. Fundamentally, STDP behaves similarly to TD learning since the temporally asymmetric learning rule binds pairs of cells if one cell spikes before (i.e. is predictive of) the other.
2. STDP can't easily learn temporally extended predictive maps but can if phase precession "compresses" input features.

Finally, we end this theoretical analysis section by examining where and why the two learning rules diverge (i.e. where STDP does not approximate TD learning). We direct the reader to studies that focus more closely on modified Hebbian learning rules to circumvent these issues, whilst pointing out that it does not have to be one or the other - the intuition for why theta phase precession helps learning applies equally well to modified learning rules which focus more closely on exactly replicating TD learning at the expense of similarity to biological STDP. We include the newly added theory section at the end of this review response document.

> *- The comparison of successor features to neural data was qualitative rather than quantitative, and often quite vague. This makes it hard to know whether the predictions of the model are actually*

consistent with real neural data. It would be much preferred if a direct quantitative comparison of the learned successor features to real data could be performed, for example, the properties of place fields near to doorways.

We agree that we could be much more specific in our comparisons to neural data, and that making quantitative comparisons to experimental recorded place cells would be a valuable contribution. To address the first point, we have clarified the presentation of our results in several places in order to make the connections to existing neural data more specific. As for making comparisons to data, we believe it is outside the scope of this work. Our primary contribution is to make quantitative comparisons between successor representations learned by TD and learned by STDP+theta. This led us to testable predictions that we have described in the discussion (page 11, paragraph beginning "*Our theory makes the prediction*") that specifically relate to the effect of impairing theta oscillations at different stages of learning (we note that these descriptions have been rewritten to be clearer in the revised manuscript). We believe that these kind of experiments would be optimal for providing datasets that would be better suited for the specific theoretical questions we are investigating here than would a post-hoc analysis of an existing datasets. Finally, we have now included theoretical analysis of the connection between STDP and TD learning (see comment above), in which readers may find a more insightful way to gain intuition about how closely this model matches SR theory and solidifies the theory contribution.

We also want to note that some prior (and in-review) work has conducted quantitative comparisons between hippocampal data and successor representations. Neuroimaging studies have shown evidence for predictive coding of spatial and non-spatial states on varying time-horizons (Garvert et al 2017, Schapiro et al 2016, Brunec and Momennejad 2022). Other studies have found that the SR did not explain under certain conditions, such as Duvelle et al 2021. de Cothi et al 2022 provide a model comparison to explain navigation behaviours in humans and rats, and found that both were best explained by a successor representation-like strategy. We also note that in recent work also under review at eLife, Ching Fang and colleagues conduct a quantitative comparison between place fields recorded from chickadees and the successor representation (Fang et al. 2022).

- Statistical structure of theta sequences. The model used by the authors is identical to that of Chadwick et al. (2015) (except for the thresholding of the Gaussian field), and so implicitly assumes that theta sequences are generated by the independent phase precession of each place cell. However, the authors mention in the introduction that other studies argue for the coordination of place cells, such that theta sequences can represent alternative futures on consecutive theta cycles (Kay et al.). This begs the question: how important is the choice of an independent phase precession model for the results of this study? For example, if the authors were to simulate a T-maze, would a model which includes cycling of alternative futures learn the successor representation better or worse than the model based on independent coding? Given that there now is a large literature exploring the coordination of theta sequences and their encoded trajectories, it would be nice to see some discussion of how the proposed mechanism depends on/relates to this.

Thank you for this suggestion. We have added a citation to Chadwick *et al*., 2015 (ref [42]) as well as the following at the beginning of the results:

"As the agent traverses the receptive field, its rate of spiking is subject to phase precession $f_{j\theta}(x,t)$ with respect to a 10 Hz theta oscillation. This is implemented by modulating the firing rate by an independent phase precession factor which varies according to the current theta phase and how far through the receptive field the agent has travelled [42] (see Methods and Fig. 1a)"

We also discuss limits of the model with regard to the Kay et al. study, as well as possible manipulations to capture this result, in a new discussion paragraph:

"*"While our model is biologically plausible in several respects, there remain a number of aspects of the biology that we do not interface with, such as different cell types, interneurons and membrane dynamics. Further, we do not consider anything beyond the most simple model of phase precession, which directly results in theta sweeps in lieu of them developing and synchronising across place cells over time [60]. Rather, our philosophy is to reconsider the most pressing issues with the standard model of predictive map learning in the context of hippocampus (e.g., the absence of dopaminergic error signals in CA1 and the inadequacy of synaptic plasticity timescales). We believe this minimalism is helpful, both for interpreting the results presented here and providing a foundation for further work to examine these biological intricacies, such as the possible effect of phase offsets in CA3, CA1 [61] and across the dorsoventral axis [62, 63], as well as whether the model's theta sweeps can alternately represent future routes [64] by the inclusion of attractor dynamics [65]."*

And elaborate on both of these points in the methods section:

"Our phase precession model is "independent" (essentially identical to Chadwick et al. (2015)[42]) in the sense that each place cell phase precesses independently from what the other place cells are doing. In this model, phase precession directly leads to theta sweeps as shown in Fig. 1. Another class of models referred to as "coordinated assembly" models [76] hypothesise that internal dynamics drive theta sweeps within each cycle because assemblies (aka place cells) dynamically excite one-another in a temporal chain. In these models theta sweeps directly lead to phase precession. Feng and colleagues draw a distinction between theta precession and theta sequence, observing that while independent theta precession is evident right away in novel environments, longer and more stereotyped theta sequences develop over time [77]. Since we are considering the effect of theta precession on the formation of place field shape, the independent model is appropriate for this setting. We believe that considering how our model might relate to the formation of theta sequences or what implications theta sequences have for this model is an exciting direction for future work."

<u>Minor comments:</u>

- When comparing the convergence rate for non-precessing vs precessing place cells, it looks as though the precessing simulation has yet to converge. What would the R2 and SNR be if the simulation were run for a longer time?

In Fig. 2 supplement 1a, we now repeat the simulation for an additional 30 minutes. In summary, convergence was approximately complete at 30 minutes (to a total of 60 minutes) and the synaptic weight matrices are not substantially different after one hour learning compared to 30 minutes. SNR continues to increase proportionally as the weights grow and so 'signal' in the weight matrices dominates over spike-time derived 'noise'.

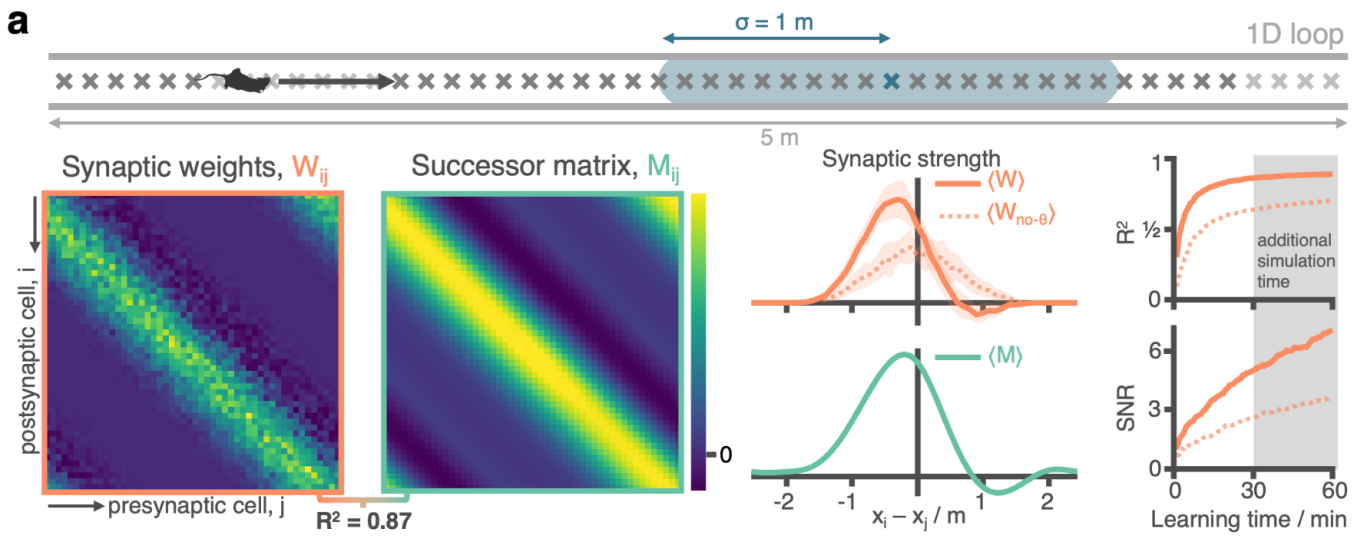


Figure 2 supplement 1: *STDP and phase precession combine to make a good approximation of the SR independent of place cell size and running speed statistics. a Figure 2 panels a-e have been repeated (additional 30 minutes simulation carried out) for ease of comparison.*

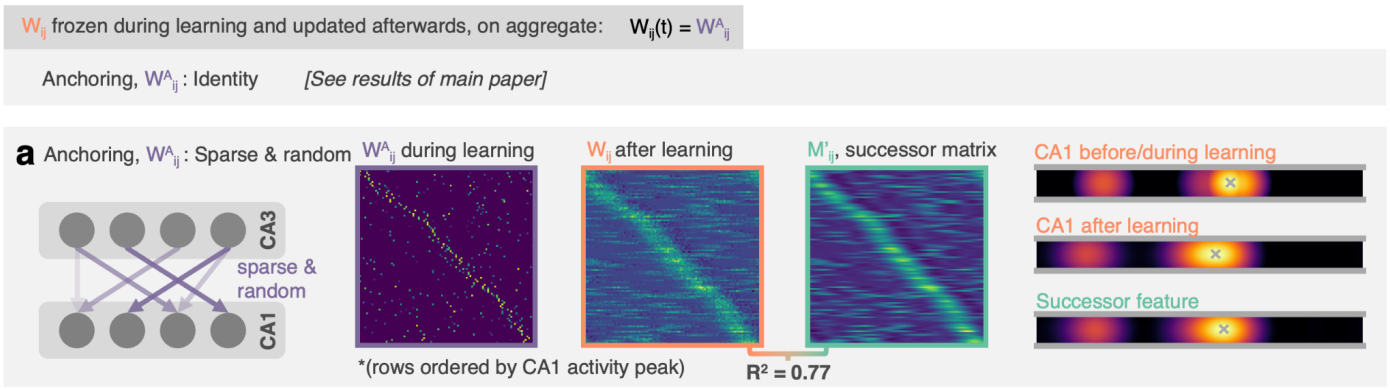
> *- The chosen peak rate of 5 Hz is lower than what is typically reported experimentally (e.g., Huxter et al., 2003). How many spikes are fires per pass in the model, and is this consistent with the experiment? What happens if firing rates are higher?*

Thank you for this comment. We experimented with higher firing rates in the original hyperparameter sweep. We observed a monotonic increase in R^2 with larger firing rate i.e. if we had chosen a higher firing rate, say 10 Hz, performance would only have been better. Instead we speculate that the energy cost of high firing rates may have played a role in preventing biology from 'optimising' this parameter.
This is stated in the results:

"We found that optimised parameters (those which result in the highest final similarity between STDP and TD weight matrices, W_{ij} and M_{ij}) were very close to the biological parameters already selected for our model from a literature search (Supp. Fig. 3c & d, parameter references also listed in figure) and, when they were used, no drastic improvement was seen in the similarity between W_{ij} and M_{ij}. The only exception was firing rate for which performance monotonically improved as it increased - something the brain likely cannot achieve due to energy constraints."

> *- Place fields show many complex properties that were not considered in the model, such as heterogeneity in size/shape, hyperdispersion in terms of their across-trial variability, banana-shaped phase precession curves, and an increase in variance from early to late in the theta cycle. While it is of course sensible for the authors to consider a simpler model which omits these phenomena, it would be nice to see some discussion and/or analysis of how these might influence the results.*

Thank you for this suggestion. In Fig. 2 supplement 2a we now explore the ability for the model to learn CA1 successor features with more complex properties such as multiple fields:



Supplementary Figure 2: *The STDP and phase precession model learns predictive maps irrespective of the weight initialisation and the weight updating schedule. In the original model weights are set to the identity before learning and kept ("anchored") there, only updated on aggregate after learning. In these panels we explore variations to this set-up. a (Left) Weights are anchored to a sparse random matrix, not the identity. (Middle) Three weight matrices show the random weights before/during learning, the weights once they have been updated on aggregate after learning and the successor matrix corresponding to the successor features of the mixed features. Matrix rows are ordered by peak CA1 activity location in order that some structure is visible. (Right) An example CA1 feature (top) before learning and (middle) after learning alongside (bottom) the corresponding successor feature.*

Furthermore, while we do not 'fully' implement the observed banana-shape of theta phase precession, we do extend the hyperparameter sweep to explore the effect of varying the proportion of the theta cycle in

which cells phase precess (i.e. the β parameter):

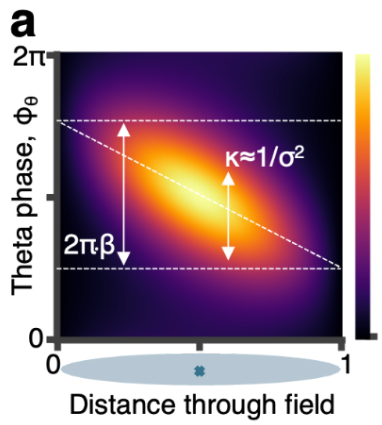


Figure 2 supplement 4: *Biological phase precession parameters are optimal for learning the SR. a We model phase precession as a von Mises centred at a preferred theta phase which precesses in time. This factor modulates the spatial firing field. It is parameterised by κ (von Mises width parameter, aka noise) and β (fraction of full 2π phase being swept, diagonal line). We showed in a previous figure that biological phase precession parameters are optimal.*

Interestingly, we find that the optimal range of β values for the STDP model to approximate TD learning coincides with those we were originally using by fitting to experimental data (specifically Jeewajee et al., 2013). One potential reason for this is that if β is too big (i.e. the precession spans the full range of the theta cycle), it is possible for cells firing late in one theta cycle to form strong acausal synaptic weights to cells firing early in the next theta sweep. We summarise the overall outcome of the updated parameter sweep in the results:

"In particular, the parameters controlling phase precession in the CA3 basis features (Fig. 2–supplement 4a) can affect the CA1 STDP successor features learnt, with 'weak' phase precession resembling learning in the absence of theta modulation (Fig. 2–supplement 4bc), biologically plausible values providing the best match to the TD successor features (Fig. 2–supplement 4d) and 'exaggerated' phase precession actually hindering learning (Fig. 2–supplement 4e; see Supplementary Materials for more details). Additionally, we find these CA1 cells go on to inherit phase precession from the CA3 population (Fig. 2–supplement 4f)."

And describe this effect in more detail in the appendices/methods:

"The optimality of biological phase precession parameters In figure 2 supplement 3 we ran a hyperparameter sweep over the two parameters associated with phase precession: κ, the von Mises parameter describing how noisy phase precession is and β, the fraction of the full 2π theta cycle phase precession crosses. The results show that for both of these parameters there is a clear "goldilocks" zone around the biologically fitted parameters we chose originally. When there is too much (large κ, large β) or too little (small κ, small β) phase precession performance is worse than at intermediate biological amounts of phase precession. Whilst – according to the central hypothesis of the paper – it makes sense that weak or non-existence phase precession hinders learning, it is initially counter intuitive that strong phase precession also hinders learning.

We speculate the reason is as follows, when β is too big phase precession spans the full range from 0 to 2π, this means it is possible for a cell firing very late in its receptive field to fire just before a cell a long distance behind it on the track firing very early in the cycle because 2π comes just before 0 on the unit circle. When κ is too big, phase precession is too clean and cells firing at opposite ends of the theta cycle will never be able to bind since their spikes will never fall within a 20 ms window of each other. We illustrate these ideas in figure 2 supplement 4 by first describing the phase precession model (panel a) then simulating spikes from 4 overlapping place cells (panel b) when phase precession is weak (panel c), intermediate/biological (panel d) and strong (panel e). We confirm these intuitions about why there exists a phase precession "goldilocks" zone by showing the weight matrix compared to the successor matrix (right hand side of panels c, d and e). Only in the intermediate case is there good similarity."

- In the simulations of rats running back and forth along a linear track, it was assumed that place maps remained fixed. However, place fields usually remap (partially or totally) when animals run along each direction of a linear track. It would be nice to see this revisited with a more realistic model which considers this remapping.

Thank you for this comment. While the primary contribution of the model and results presented here pertain more to the role of theta compression (and STDP) in approximating TD learning in the connections between CA3-CA1, we agree that understanding remapping characteristics, such as directionality on a linear track, is a topic of huge importance. While the emergence of this directionality is proportionally greater in CA1 than CA3, it remains true that this directionality is already present in CA3 (McNaughton, Barnes & O'Keefe 1983). As such, it seems likely that at least some of the mechanisms causing such remapping lay upstream of the CA3-CA1 connections we focus on here. Nonetheless, the STDP-SR learning mechanism presented here would remain unaltered by imposing such directionality in the CA3 basis features, where the distinction between learning on directional or allocentric basis features is functionally identical to our simulations of the agent agent travelling along the track in only one direction (i.e. Fig. 2a-e) or both directions (i.e. Fig. 2f-j) respectively.

- The model for theta sequence generation is essentially identical to that of Chadwick et al. (2015). This should be cited and the relationship discussed.

Thank you for highlighting this, we have added this citation ([42]) to the main text at the beginning of the results:

"As the agent traverses the receptive field, its rate of spiking is subject to phase precession $f_{j\theta}(x,t)$ with respect to a 10 Hz theta oscillation. This is implemented by modulating the firing rate by an independent phase precession factor which varies according to the current theta phase and how far through the receptive field the agent has travelled [42] (see Methods and Fig. 1a)"

And further describe this similarity in the methods section:

"Our phase precession model is "independent" (essentially identical to Chadwick et al. (2015)[42]) in the sense that each place cell phase precesses independently from what the other place cells are doing. In this model, phase precession directly leads to theta sweeps as shown in Fig. 1. Another class of models referred to as "coordinated assembly" models [76] hypothesise that internal dynamics drive theta sweeps within each cycle because assemblies (aka place cells) dynamically excite one-another in a temporal chain. In these models theta sweeps directly lead to phase precession. Feng and colleagues draw a distinction between theta precession and theta sequence, observing that while independent theta precession is evident right away in novel environments, longer and more stereotyped theta sequences develop over time [77]. Since we are considering the effect of theta precession on the formation of place field shape, the independent model is appropriate for this setting. We believe that considering how our model might relate to the formation of theta sequences or what implications theta sequences have for this model is an exciting direction for future work."

- I struggled to understand Figure 1b - do the concentric circles represent firing rate? If so, perhaps this can be labelled. And similarly for the arrows.

Thank you for highlighting this, we have added a key to clarify.

- Typo: "a phenomena".

Thank you for raising this, it has been corrected.

- Typo: "in observed in CA1 place cells than CA3 place cells".

This has been corrected, thank you.

Finally we duplicate for convenience the new theory section which can be found in the appendices/methods of the manuscript.

5.8 A theoretical connection between STDP and TD learning

Why does STDP between phase precessing place cells approximate TD learning? In this section we attempt to shed some light on this question by analytically studying the equations of TD learning. Ultimately, comparisons between these learning rules are difficult since the former is inherently a discrete learning rule acting on pairs of spikes whereas the latter is a continuous learning rule acting on firing rates. Nonetheless, in the end we will draw the following conclusions:

1. *In the first part we will show that, under a small set of biologically feasible assumptions, temporal difference learning "looks like" a spike-time dependent temporally-asymmetric Hebbian learning rule (that is, roughly, STDP) where the temporal discount time horizon, τ is equal to the synaptic plasticity timescale O(20 ms).*
2. *2. In the second part we will see that this limitation that the temporal discount time horizon is restricted to the timescale of synaptic plasticity (i.e. very short) can be overcome by compressing the inputs. Phase precession, or more formally, theta sweeps, perform exactly the required compression.*

In sum, there is a deep connection between TD learning and STDP and the role of phase precession is to compress the inputs such that a very short predictive time horizon amounts to a long predictive time horizon in decompressed time coordinates. We will finish by discussing where these learning rules diverge and the consequences of their differences on the learned representations. The goal here is not to derive a mathematically rigorous link between STDP and TD learning but to show that a connection exists between them and to point the reader to further resources if they wish to learn more.

5.8.1 Reformulating TD learning to look like STDP

First, recall that the temporal difference (TD) rule for learning the successor features $\psi_i(x)$ defined in Eqn. (19) takes the form:

$$\frac{d\mathbf{M}_{ij}}{dt} = \eta \delta_i(t) \mathbf{e}_j(t) \tag{30}$$

where M_{ij} are the weights of the linear function approximator, Eqn. (3) and $\delta_i(t)$ is the continuous temporal difference error defined in Eqn. (24). $e_j(t)$ is the eligibility trace for feature j defined according to

$$\mathbf{e}_j(t) = \int_{-\infty}^{t} \frac{1}{\tau_\mathbf{e}} e^{\frac{t-t'}{\tau_\mathbf{e}}} f_j(\mathbf{x}(t'))dt' \tag{31}$$

or, equivalently, by its dynamics (which we will make use of)

$$\mathbf{e}_j(t) = f_j(t) - \tau_\mathbf{e} \dot{\mathbf{e}}_j(t). \tag{32}$$

where $\tau_e \in [0,\tau]$ is a 'free' parameter, the eligibility trace timescale, analogous to λ in discrete TD(λ). When $\tau_e = 0$ we recover the learning rule we use to learn successor features, "TD(0)", in Eqn. (21).
Subbing Eqn. (24) and Eqn. (32) into this update rule, Eqn. (30), rearranges to give

$$\frac{d\mathbf{M}_{ij}}{dt} = \eta\big(f_i \mathbf{e}_j - \psi_i f_j + \tau \dot{\psi}_i \mathbf{e}_j - \tau_\mathbf{e} \psi_i \dot{\mathbf{e}}_j\big) \tag{33}$$

where we redefined $\eta \leftarrow \eta' = \eta/\tau$. Now let the predictive time horizon be equal to the eligibility trace timescale. This setting is also called TD(1) or Monte Carlo learning,

$$\tau = \tau_\mathbf{e} \tag{34}$$

Now

$$\frac{d\mathbf{M}_{ij}}{dt} = \eta\big(f_i \mathbf{e}_j - \psi_i f_j + \tau_\mathbf{e} \frac{d}{dt}(\psi_i \mathbf{e}_j)\big). \tag{35}$$

The final term in this update rule, the total derivative, can be ignored with respect to the stationary point of the learning process. To see why, consider the simple case of a periodic environment which repeats over a time period T – this is

true for the 1D experiments studied here. Learning is at a stationary point when the integrated changes in the weights vanish over one whole period:

$$0 = \int_t^{t+T} dt' \dot{\mathsf{M}}_{ij}(t') = \eta \int_t^{t+T} dt' \big(f_i \mathsf{e}_j - \psi_i f_j\big) + \eta\tau_{\mathsf{e}} \int_t^{t+T} dt' \frac{d}{dt'}(\psi_i(t')\mathsf{e}_j(t')) \tag{36}$$

$$= \eta \int_t^{t+T} dt' \big(f_i \mathsf{e}_j - \psi_i f_j\big) + \eta\tau_{\mathsf{e}} \big[\psi_i(t+T)\mathsf{e}_j(t+T) - \psi_i(t)\mathsf{e}_j(t)\big] \tag{37}$$

$$= \eta \int_t^{t+T} dt' \big(f_i \mathsf{e}_j - \psi_i f_j\big) \tag{38}$$

where the last term vanishes due to the periodicity. This shows that the learning rule converges to the same fixed point (i.e. the successor feature) irrespective of whether this term is present and it can therefore be removed. The dynamics of this updated learning rule won't strictly follow the same trajectory as TD learning but they will converge to the same point. Although strictly we only showed this to be true in the artificially simple setting of a periodic environment it is more generally true in a stochastic environment where the feature inputs depend on a stationary latent Markov chain[43].
Thus a valid learning rule which converges onto the successor feature can be written as

$$\frac{d\mathsf{M}_{ij}}{dt} = \eta\big(f_i(t)\mathsf{e}_j(t) - \psi_i(t)f_j(t)\big) \tag{39}$$

Claim: this looks like a continuous analog of STDP acting on the weights between a set of input features, indexed j, and a set of downstream "successor features" indexed i. Each term in the above learning rule can be non-rigorously identified as follows, a key change is that the successor features neurons have two-compartments; a somatic compartment and a dendritic compartment:

• f(t) := $V^{soma}(t)$ is the somatic membrane voltage which is primarily set by a "target signal". In general ii this target signal could be any reward density function, here it is the firing rate of the ith input feature.
• ψ(t) := $V^{dend}(t)$ is the voltage inside a dendritic compartment which is a weighted linear sum of the input currents, Eqn. (3). This compartment is responsible for learning the successor feature by adjusting its input weights, Mij, according to equation (39).
• f(t) := I(t) are the synaptic currents into the dendritic compartment from the upstream features.
• e(t) := ˜I (t) are the low-pass filtered eligibility traces of the synaptic input currents.

$$\frac{d\mathsf{M}_{ij}}{dt} = \eta\Big(\underbrace{\mathsf{V}_i^{\mathrm{soma}}(t)\tilde{\mathsf{I}}_j(t)}_{\text{pre-before-post potentiation}} - \underbrace{\mathsf{V}_i^{\mathrm{dend}}(t)\mathsf{I}_j(t)}_{\text{post-before-pre depression}}\Big) \tag{40}$$

This learning rule, mapped onto the synaptic inputs and voltages of a two-compartment neuron, is Hebbian. The first term potentiates the synapse M_{ij} if there is a correlation between the low-pass filtered presynaptic current and the somatic voltage (which drives postsynaptic activity). More specifically this potentiation is is temporally asymmetric due to the second term which sets a threshold. A postsynaptic spike (e.g. when $V_i^{soma}(t)$ reaches threshold) will cause potentiation if

$$\mathsf{V}_i^{\mathrm{soma}}(t)\tilde{\mathsf{I}}_j(t) > \mathsf{V}_i^{\mathrm{dend}}(t)\mathsf{I}_j(t) \tag{41}$$

but since the eligibility trace decays uniformly after a presynaptic input this will only be true if the postsynaptic spike arrives very soon after. This is pre-before-post potentiation. Conversely an unpaired presynaptic input (e.g. when Ij(t) spikes) will likely cause depression since this bolsters the second depressive term of the learning rule but not the first (note this is true if its synaptic weight is positive such that $V^{dend}(t)$ will be high too). This is analogous to post-before-pre depression. Whilst not identical, it is clear this rule bears the key hallmarks of the STDP learning rule used in this study, specifically: pre-before-post synaptic activity potentiates a synapse if post synaptic activity arrive within a short time of the presynaptic activity and, secondly, post-before-pre synaptic activity will typically result in depression of the synapse.
Intuitively it now makes sense why asymmetric STDP learns successor features. If a postsynaptic spike from the ith neuron arrives just after a presynaptic spike from the jth feature it means, in all probability, that the presynaptic input features is "predictive" of whatever caused the postsynaptic spike which in this case is the ith feature. Thus if we want to learn a function which is predictive of the ith features future activity (its successor feature) we should increase the synaptic weight M_{ij}. Finally, identifying that this learning rule looks similar to STDP fixes the timescale of the eligibility

trace to be the timescale of STDP plasticity i.e. O(20 − 50 ms). And to derive this learning rule we required that the temporal discount time horizon must equal the eligibility trace timescale, altogether:

$$\tau = \tau_{\mathbf{e}} = \tau_{\mathrm{STDP}} \approx 20 - 50 \text{ ms} \tag{42}$$

This limits the predictive time horizon of the learnt successor feature to a rather useless – but importantly non-zero – 20-50 ms. In the next section we will show how phase precession presents a novel solution to this problem.

5.8.2 Theta phase precession compresses the temporal structure of input features

We showed in Fig. 1 how phase precession leads to theta sweeps. These phenomena are two sides of the same coin. Here we will start by positing the existence of theta sweeps and show that this leads to a potentially large amount of compression of the feature basis set in time.
First, consider two different definitions of position. x_T (t) is the "True" position of the agent representing where it is in the environment at time t. x_E(t) is the "Encoded" position of the agent which determines the firing rate of place cells which have spatial receptive fields fi(x_E(t)). During a theta sweep the encoded position x_E (t) moves with respect to the true position x_T(t) at a relative speed of v_S (t) where the subscript S distinguishes the "Sweep" speed from the absolute speed of the agent $\dot{x}_T$(t) = v_A(t). In total, accounting for the motion of the agent:

$$\dot{\mathbf{x}}_E(t) = \mathbf{v}_A(t) + \mathbf{v}_S(t) \tag{43}$$

Now consider how the population activity vector changes in time

$$\frac{d}{dt} f_i^T(\mathbf{x}_E(t)) = \nabla_{\mathbf{x}} f_i^T(\mathbf{x}) \cdot \dot{\mathbf{x}}_E(t) = \nabla_{\mathbf{x}} f_i^T(\mathbf{x}) \cdot (\mathbf{v}_A(t) + \mathbf{v}_S(t)) \tag{44}$$

and compare the time how it would varying in time if there was no theta sweep (i.e x_E(t) = x_T(t))

$$\frac{df_i^T(\mathbf{x}_T(t))}{dt} = \nabla_{\mathbf{x}} f_i^T(\mathbf{x}) \cdot \frac{d\mathbf{x}_T(t)}{dt} = \nabla_{\mathbf{x}} f_i^T(\mathbf{x}) \cdot \mathbf{v}_A(t). \tag{45}$$

They are proportional. Specifically in 1D, where the sweep is observed to move in the same direction as the agent (from behind it to in front of it) this amount to compression of the temporal dynamics by a factor of

This "compression" is also true in 2D where sweeps are also observed to move largely in the same direction as the agent.
If this compression is large it would solve the timescale problem described above. This is because learning a successor feature with a very small time horizon, τ, where the input trajectory is heavily compressed in time by a factor of κ_θ amounts to the same thing as learning a successor feature with a long time horizon τ' = τκ_θ where the inputs are not compressed in time.
What is v_S, and is it fast enough to provide enough compression to learn temporally extended SRs? We can make a very rough ballpark estimate. Data is hard to come by but studies suggest the intrinsic speed of theta sweeps can be quite fast. Figures in Feng et al. (2015), Wang et al. (2020) and Bush et al. (2022) show sweeps moving at up to, respectively, 9.4 ms^{-1}, 8.5 ms^{-1} and 2.3 ms^{-1}. A conservative range estimate of $v_S \approx 5 \pm 5$ ms^{-1} accounts for very fast and very slow sweeps. The timescale of STDP is debated but a reasonable conservative estimate would be around $\tau^{STDP} \approx 35 \pm 15 \times 10^{-3}$ s which would cover the range of STDP timescales we use here. The typical speed of a rat, though highly variable, is somewhere in the range $v_A \approx 0.15 \pm 0.15$ ms^{-1}. Combining these (with correct error analysis, assuming Gaussian uncertainties) gives an effective timescale increase of

$$\tau' = \tau k_\theta = \tau_{\mathrm{STDP}} \frac{v_A + v_S}{v_A} \approx 1.1 \pm 1.7\text{s} \tag{47}$$

Therefore we conclude theta sweeps can provide enough compression to lift the timescale of the SR being learn by STDP from short synaptic timescales to relevant behavioural timescales on the order of seconds. Note this ballpark estimate is not intended to be precise, and doesn't account for many unknowns for example the covariability of sweep speed with running speed[cite], variability of sweep speed with track length[cite] or cell size[cite] which could potentially extend this range further.

5.8.3 Differences between STDP and TD learning: where our model doesn't work

We only drew a hand-waving connection between the TD-derived Hebbian learning rule in Eqn. (39) and STDP. There are numerous difference between STDP and TD learning, these include the fact that

1. *Depression in Eqn. (39) is dependent on the dendritic voltage which is not true for our STDP rule.*

2. *Depression in Eqn. (39) is not explicitly dependent on the time between post and presynaptic activity, unlike STDP.*
3. *Eqn. (39) is a continuous learning rule for continuous firing rates, STDP is a discrete learning rule applicable only to spike trains.*

Analytic comparison is difficult due to this final difference which is why in this paper we instead opted for empirical comparison. Our goal was never to derive a spike-time dependent synaptic learning rule which replicates TD learning, other papers have done work in this direction (see [43, 41]), rather we wanted to (i) see whether unmodified learning rules measured to be used by hippocampal neurons perform and (ii) study whether phase precession aids learning. Under regimes tested here, STDP seems to hold up well.

These differences aside, the learning rule does share other similarities to our model set-up. A special feature of this learning rule is that it postulates that somatic voltage driving postsynaptic activity during learning isn't affected by the neurons own dendritic voltage. Rather, dendritic voltages affect the plasticity by setting the potentiation threshold. These learning rules have been studies under the collective name of "voltage dependent" Hebbian learning rules[CITE]. This matches the learning setting we use here where, during learning, CA1 neurons are driven by one and only one CA3 feature (the "target feature") whilst the weights being trained Wij don't immediately effect somatic activity during learning. The lack of online updating matches the electrophysiological observation that plasticity between CA3 and CA1 is highest during the phase of theta when CA1 is driven by entorhinal cortex and lowest at the phase when CA3 actually drives CA1[83].

Finally, there is one clear failure for our STDP model – learning very long timescale successor features. Unlike TD learning which can 'bootstrap' long timescale associations through intermediate connections, this is not possible with our STDP rule in its current form. Brea et al. (2016)[43] and Bono et al. (2022)[41] show how Eqn. (39) can be modified to allow long timescale SRs whilst still enforcing the timescale constraint we imposed in Eqn. (34) thus still maintaining the biological plausibility of the learning rule, this requires allowing the dendritic voltage to modify the somatic voltage during learning in a manner highly similar to bootstrapping in RL. Specifically in the former study this is done by a direct extension to the two-compartment model, in the latter it is recast in a one-compartment model although the underlying mathematics shares many similarities. Ultimately both mechanisms could be at play; even in neurons endowed with the ability to bootstrap long timescale association with short timescale plasticity kernels phase precession would still increase learning speed significantly by reducing the amount of bootstrapping required by a factor of κ_θ, something we intend to study more in future work. Finally it isn't clear what timescales predictive encoding in the hippocampus reach, there is likely to be an upper limit on the utility of such predictive representations beyond which the animal use model-based methods to find optimal solution which guide behaviour.