We thank the reviewers for their thoughtful reviews. Below are specific comments to each reviewer.

Reviewer 1: We briefly discussed a suggestion for biological experiments to test the theory in line 428. We could expand on this. If we measured pre and post-synaptic spike trains, and also recorded changes in post-synaptic potentials to measure changes in synaptic weight, we could use hidden Markov model techniques to find the best-fit synaptic model. Then given our theory, we could match this measured synaptic model to optimal models to infer which timescales the synapse operates on. Moreover, it is important to note that the modeling framework we have used to model synapses has been previously published in journals like Neuron and Nature Neuroscience, so the modeling framework has been vetted for sufficient biological realism. However, nobody has previously provided a general theory of the incredibly flexible space of synaptic models achievable within this framework, and that is our novel contribution. Regarding the discrepancy between the envelope and the numerical results: the envelope is just an upper bound and we do not claim that it is a tight bound at all times (see discussion). The reason is that equation (18) is not a complete set of constraints, as discussed below it. In the paragraph from line 377 to 400, we discuss this point further. Note that the numerical results are not always instructive as the numerical procedures can be prevented from reaching the global maximum by local maxima. This is shown by the fact that our hand designed models can outperform the numerical methods at late times. In fact, the apparent drop-off of the solid red curve at very late times comes from not allowing small enough epsilon. We will fix this in fig5. Yes, the level of noise tolerance is an important consideration. Note that the stochasticity of the models is an expression of biological noise. In fact, the type of noise described by the reviewer could be included by adding extra states to the model with the different transition rates. However, the synapses should not be allowed to optimize all of this noise away. It would be interesting to consider such limits on the noise levels. I think this would be beyond the scope of this work, but should be added to the discussion section. Reviewer 2: We'll add a derivation of equation 3 to the supplement and main text. We could add a description in the main text along the lines of: "The factor of p^infinity describes the synapses being in the steady-state distribution before the memory is encoded. The factor of (M^pot-M^dep) comes from the encoding of the memory at t=0, with w\_ideal being +/-1 in synapses that are potentiated/depotentiated. The factor of exp(rt W^F) describes the subsequent evolution of the probability distribution, averaged over all sequences of plasticity events and the factor of w indicates the readout via the synaptic weight." We referred to all of the original parts of the supplement in the main paper, but we can be more explicit, and would be happy to modify the paper to do so. Reviewer 3: line 195: in the introduction, we cited [17], which describes diversity in synaptic structure across the vertebrate brain. This could be related to optimization for different timescales, but anything more than speculation would require a better understanding of the relation between structure and function (which is what we intend to begin with this work). line 220: Yes, fig2a,b is an example of why it is difficult to map molecular states to functional states as, despite appearances, these models actually only have two functional states due to their equivalence to fig 2c. One experimental investigation of this could be along the lines of the experiment described in line 428 (see first paragraph of our reply to reviewer 1 above for elaboration). Presumably we will find fewer functional states than molecular states. Making the link between molecular and functional states is an important research question for neurobiology. Our work helps make progress towards this by providing a theory for how functional states might be related to each other when memory is optimal, giving experimentalists clues to look for. line 294: agreed. line 299: The fact that eta\_i depends on M^pot/dep is taken into account. In equation (57) of the supplement, the term involving c\_g comes from this dependence. If the reviewer is concerned that the order of the eta\_i could change during the maximization procedure: note that necessary conditions for a maximum only require that there is no infinitesimal perturbation that increases the area. Therefore we need only consider an infinitesimal neighborhood of the model, in which the order will not change. line 304: Yes. We'll add this phrase to the text. line 326: In fig2(a,b), we need not have an equal number of w=+/-1 states (we could change the figure to reflect this), but those states are not functionally relevant, as shown by their equivalence to fig2c. The model in fig2c, of course, has no room for such asymmetry. The models in figs4,5b do need to have equal numbers of +/- states, asymmetry would make them worse. For fig4, the effect of the asymmetry would be reduced in the limit as epsilon -> 0, as all states other than the end states would have very small p^infinity. It is important to note that while our figures do show symmetric models, our proofs are general and apply to asymmetric models as well. line 415: It is only for t\_0 > sqrt(N)M that the models that nearly touch the envelope are linear chains. The model in fig5b is not a linear chain, as it has shortcut transitions, but they are only the best models (that we have found) for times t\_0 < sqrt(N)M. Would it help if we repeated fig4 as fig5c?