We thank the reviewers for their constructive feedback.

Reviewer 2 pointed out that cross frequency coupling is “speculative” but this topic has been discussed widely in particular in Parkinson which is cited also in this paper. There is also a very recent modelling paper:

> Belić JJ, Halje P, Richter U, Petersson P and Hellgren Kotaleski J (2016) Untangling Cortico-Striatal Connectivity and Cross-Frequency Coupling in L-DOPA-Induced Dyskinesia. Front. Syst. Neurosci. 10:26. doi: 10.3389/fnsys.2016.00026

However, we agree with the reviewer that more neurophysiologcal work needs to be done to explore cross frequency coupling and plasticity. Clearly the brain uses both high frequency bands which can easily change plasticity and low frequency bands which cannot.

Reviewer 1/3: H_0 is simply the transfer function of the fixed feedback loop – e.g. if it were a simple thermostat trying to maintain a desired temperature, it would be a threshold function that mapped temperatures onto a binary output. P_1 is the function that transforms the actions back into sensory inputs V_i used for learning. Happy to add this to the text.

Reviewers 1/3: we have confused you with the unexplained term ’reflex’, apologies! In figure 1, we see that there is a fixed feedback loop with setpoint SP; this is what we are calling a ’reflex’, as an analogy with the reflex circuits that are often seen in biology. The reflex for the shooter game is rather artificial, but reflexes are common biological control mechanisms for coping with disturbances. However they have the drawback that they are purely reactive. The contribution here is:

- To show that the reflex can also provide a learning signal for training a deep network, and that network is then learning input control rather than output control - it is learning to keep the reflex silent rather than produce target outputs.
- To achieve learning where the errors are defined at the inputs, we develop a learning rule is not gradient-based, but rather where both activity and error signals are propagated forwards in the same weighted fashion. The learning rule is then correlation-based between these two signals.
- The “reflex” was chosen because it’s possible to treat it analytically with control theory. However, the error can also originate from, for example the “critic” of TD learning aka the “reward prediction error” (and is thus also biologically realistic). This would indeed allow more predictive tasks such as Atari, etc but is far beyond the scope to show a very new concept as pointed out by all reviewers.

All reviewers, in particular reviewer 1 about Figure 2B: We thought that diagram would help to understand the learning Eq. 2&3, where we correlate the activity v_j at neuron j with the error signal e_k at neuron k. This is identical to backpropagation of course - the key difference lies in the fact that the error is calculated at the inputs and propagated forwards in a weighted fashion (Eq.3), as is the case in closed loop systems. We’d be very happy to expand this section to improve the clarity of the algorithm as suggested by reviewer 1 and tie this in with error feedback.

Reviewer 1 about behavioural flexibility: The reason we would expect more behavioural flexibility from this approach is that learning does not explicitly evaluate the network outputs at all. Of course, the inputs do relate to the outputs, but only indirectly, and only via the transfer function $P_{0/1}$ of the environment. So the outputs are less constrained than they would be under Q-learning for example, where the Q value is a direct function of the outputs (and the world state). We should probably say “in principle this leads to greater flexibility”, until we show evidence of this.

Reviewer 3: The goal was to show the feasibility of an algorithm that can function in an arbitrarily deep network. It is true that the actual network used was not particularly deep however. In section 4, we say "let us assume two layers j and k", but this is merely to illustrate how the learning in one layer relates to its neighbouring layer to demonstrate a novel concept.

Reviewer 3: For the shooter example, unfortunately some details have been lost during editing. There are 3 output neurons with $tanh$ activation; a negative output means ‘rotate left’, a positive ‘rotate right’. The 3 neurons operate with different gains, to give finer control. The terms g_{net} and g_{err} are the gains for the reflex, and the learned control signal. Momentum was 0.5; weights initialiased in a zero-mean uniform distribution, scaled by the number of weights outputting each neuron.

Reviewer 1, Equation 4: The term ρ is the gain of the fixed feedback loop (the reflex), and can be absorbed into H_0.

Reviewer 1, Equation 5:

$$\Delta w_{ij} = X_0(z)V_i(-z) \tag{1}$$

is simply a correlation of X_0 with V_i, and . The term $(-z)$ appears because correlation involves reversing one of the signals in time.