Deep learning without weight transport

Mohamed Akrout[1], Collin Wilson[1], Peter C. Humphreys[2], Timothy Lillicrap[2,3], Douglas Tweed[1,4]

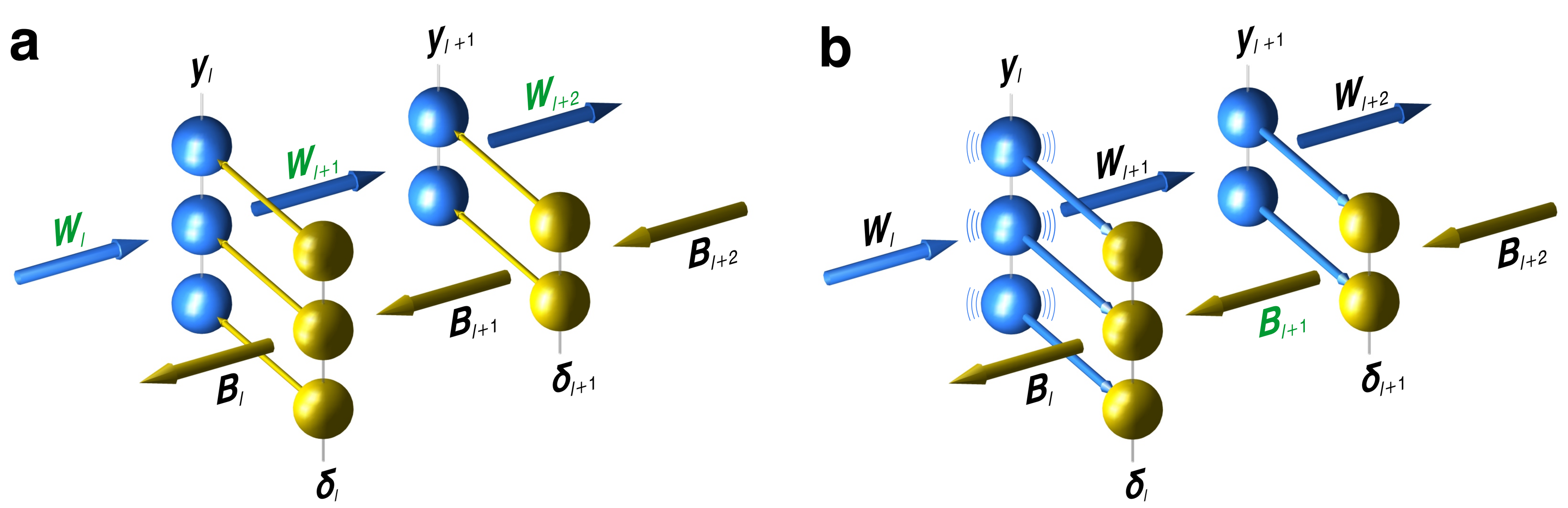
[1]University of Toronto; [2]DeepMind, [3]University College London, [4]York University

Aim

- How do neural networks in the brain learn sensorimotor tasks?
- In artificial networks, the most effective learning mechanism is *backprop*, used for *deep learning*, but it can't work in the brain because it relies on *weight transport*, where synapses interact in ways that aren't possible biologically.
- We describe two new mechanisms that learn nearly as well as backprop but without weight transport.

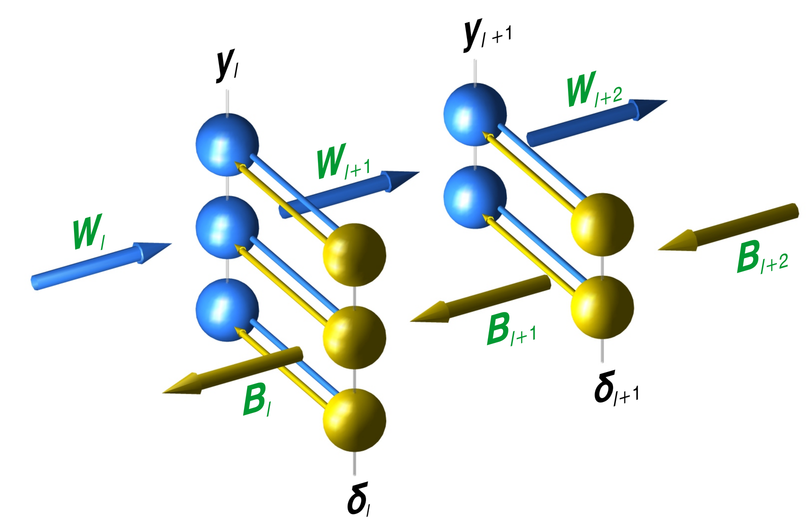
Weight mirrors

- In backprop (panel **a** below), feedback neurons (gold) send teaching signals $\boldsymbol{\delta}$ to forward neurons (blue).
- Synapses in the feedback path must have the same weights (strengths) as those in the forward path, i.e. feedback weight matrices $\boldsymbol{B}$ must be transposes of forward-path matrices $\boldsymbol{W}$.
- In the brain, how could synapses coordinate themselves in this way? That is the *weight transport problem*.



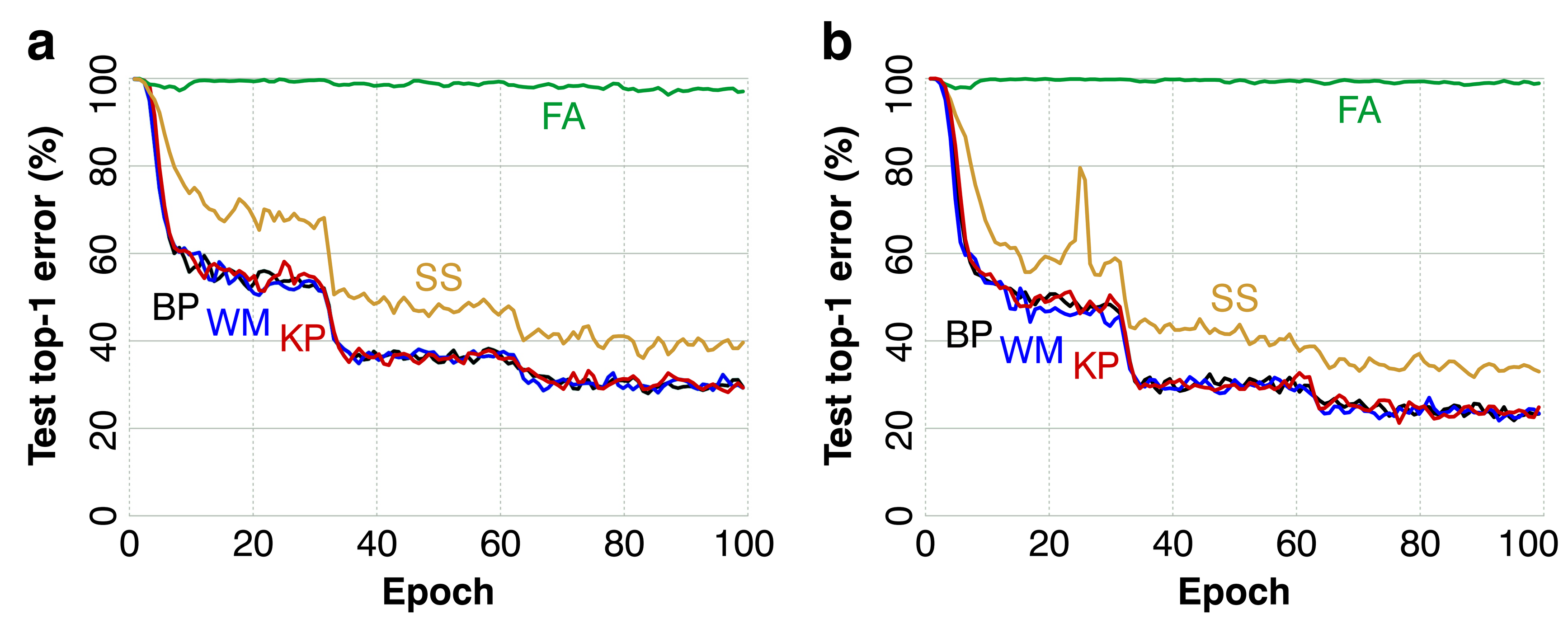
- We propose that the network enters a *mirror mode* (panel **b**), where some forward neurons fire noisily, and forward cells drive feedback cells, so $\boldsymbol{\delta} = \boldsymbol{y}$.
- A Hebbian learning rule, $\Delta \boldsymbol{B}_{l+1} = \eta\, \boldsymbol{\delta}_l \boldsymbol{\delta}_{l+1}^{\mathsf{T}} - \lambda \boldsymbol{B}_{l+1}$, where η and λ are positive constants, drives $\boldsymbol{B}_{l+1}$ to become a positive-definite multiple of $\boldsymbol{W}_{l+1}^{\mathsf{T}}$.
- This circuit, the *weight mirror*, learns without sensory input, so it could tune feedback paths in sleep or in utero.

Network for Kolen-Pollack learning

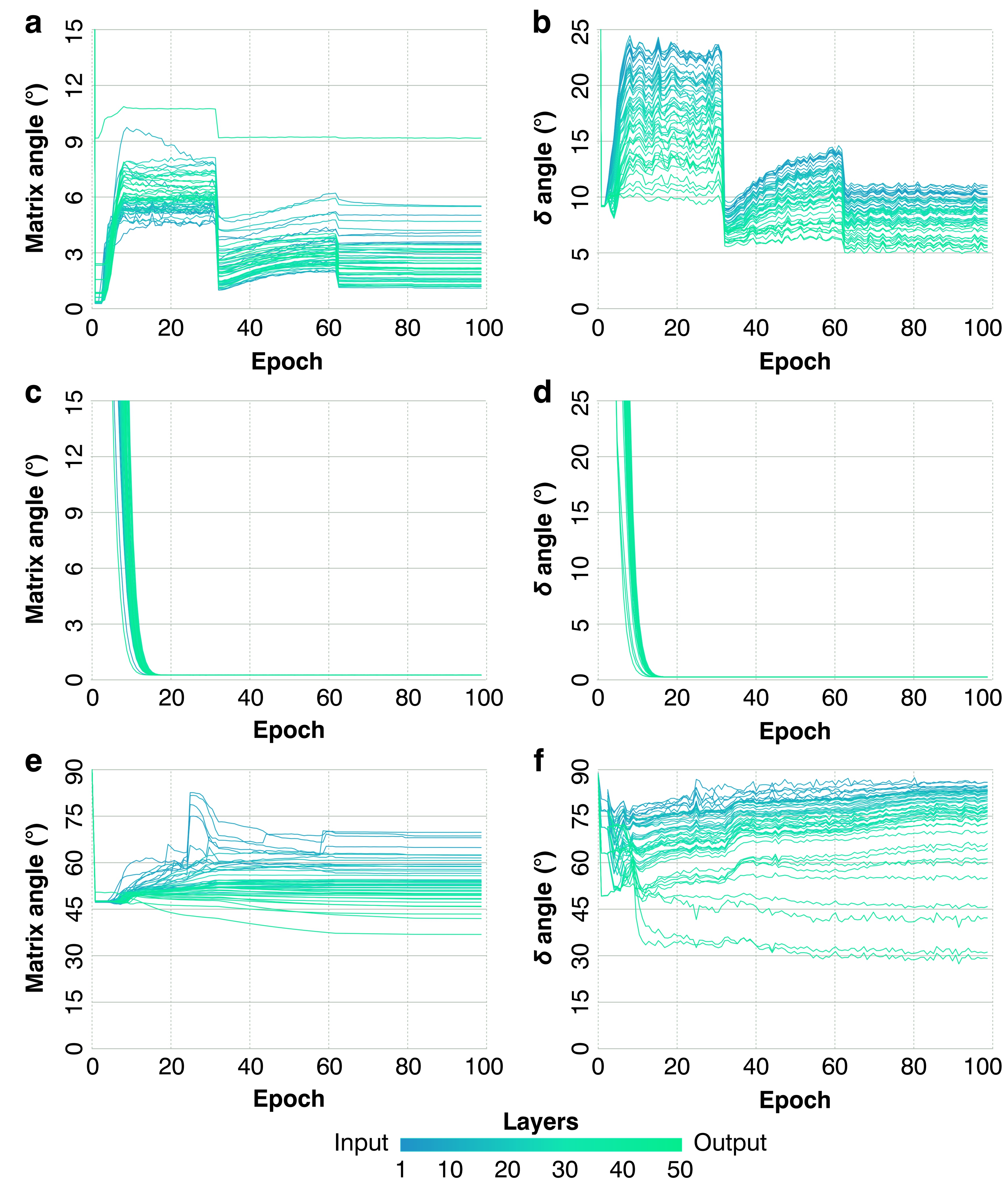


- Kolen and Pollack (1994) described an algorithm that made $\boldsymbol{B} = \boldsymbol{W}^{\mathsf{T}}$, but it worked by transporting weight-*changes*, which is also unbiological.
- We propose the above circuit and a feedback-path learning rule $\Delta \boldsymbol{B}_{l+1} = -\eta\, \boldsymbol{y}_l \boldsymbol{\delta}_{l+1}^{\mathsf{T}} - \lambda \boldsymbol{B}_{l+1}$, which together with the backprop rule $\Delta \boldsymbol{W}_{l+1} = -\eta\, \boldsymbol{\delta}_{l+1} \boldsymbol{y}_l^{\mathsf{T}} - \lambda \boldsymbol{W}_{l+1}$ implement Kolen-Pollack without transmitting any synaptic data.

Experiments



- We tested our weight mirrors (WM) and Kolen-Pollack network (KP) against other proposed mechanisms for biological learning — *feedback alignment* (FA, Lillicrap et al. 2016) and *sign-symmetry* (SS, Xiao et al. 2018) — on the ImageNet visual recognition task.
- WM and KP outperformed FA and SS, and kept pace with backprop (BP), when run on either the ResNet-18 network (panel **a**) or the deeper ResNet-50 (panel **b**).



- WM (panels **a**, **b**) and KP (**c**, **d**) kept $\boldsymbol{B}$ matrices and $\boldsymbol{\delta}$ vectors close to what BP would have done, but SS (**e**, **f**) did not.

Conclusions

- Weight mirrors and the Kolen-Pollack circuit show how neural networks can achieve near-backprop-level learning without the unbiological device of weight transport.