Decrease in neuron activation after ablating attention head 1.2 Decrease in neuron activation 1.0 0 0000 8.0 0.6 8 0.4 0.2 0.0 Attn Head Active Attn Head Inactive Overall