



**Figure R1: Results for batch size = 1.** Each panel shows the performance of ResNet-50 in Experiments 1 (Patch) and 2 (Segment) under the four test conditions (Both, Conflict, Shape, Non-shape). Unlike other simulations reported in this study, these results were obtained for a pure SGD setting, with a batch size of 1. The larger error bars are a consequence of smaller number of training epochs and seeds as learning takes much longer with a small batch. Critically, we obtain the same pattern as the main experiments: changing the condition from Both to Conflict shows a large drop in performance. When the non-shape feature is more predictive than shape (Figure b), we obtain the same high–low–low–high pattern that indicates the network learns almost entirely based on the non-shape feature.