Supplementary Figure 2: Inhibition in a deep environment. The outcomes \( O \) are approached by sequentially walking through \( K = 4 \) levels. Only \( T^4 \) states lead to outcomes. (A,D): True values without inhibition are shown by black line. It is constant for each level and valence as, or illustration, all outcomes were assigned the same positive value (+1 or -1). The reward of the states \( T \) is zero and shown by the dash-dotted line. The grey point display the estimated values of the states under inhibition \( \alpha_{5HT} = 20 \). There is a positive bias in all states, but it is more pronounced in the states with true negative values. In (D), the dash-dotted line indicates that states \( T^4 \) now carry reward -0.4, while states \( T^k \) for \( k = \{1, 2, 3\} \) now have true negative values and \( T^k \) for \( k = \{1, 2, 3\} \) have true positive values. (B,E): Probabilities of ending thought sequence in \( O^+ \) or \( O^- \). (C,F): Effect of preferentially choosing actions according to their valence on the average value of states. The arrow indicates increasing \( \theta \). In (C), larger \( \theta \) are advantageous, in (F), smaller \( \theta \) are better.