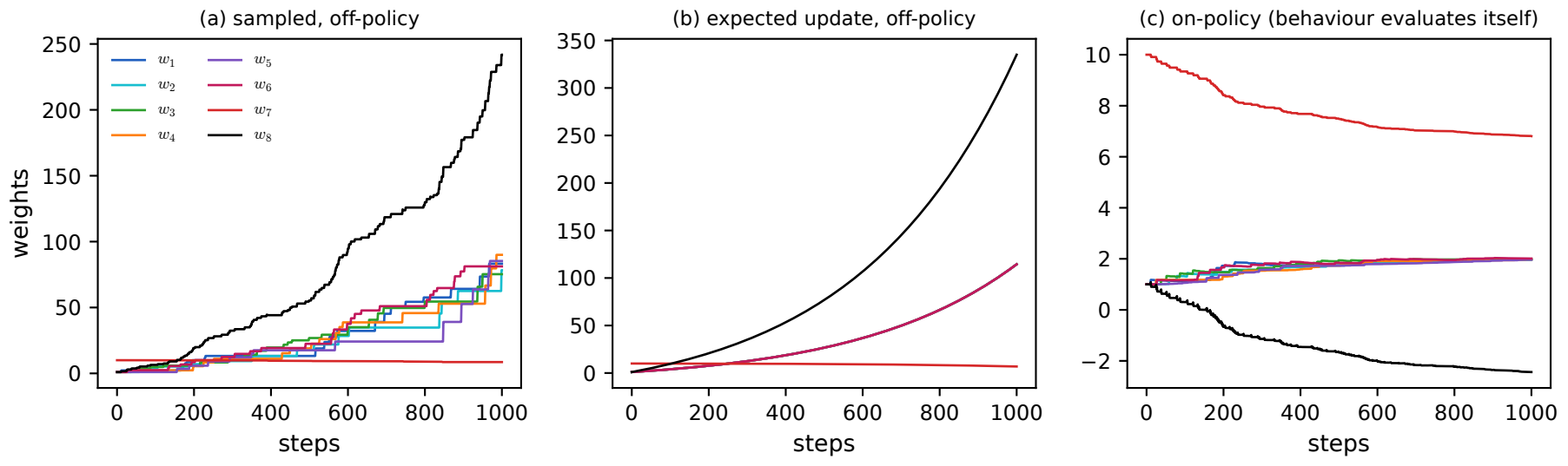

(a) sampled, off-policy
(b) expected update, off-policy
(c) on-policy (behaviour evaluates itself)
weights
steps
steps
steps
w_1
w_2
w_3
w_4
w_5
w_6
w_7
w_8