Consensus (N = 2) MDP Q-value, property steps_max, deep 5-layered network

Legend:
- One-hot encoding
- Ordinal encoding
- One-hot encoding (Q-table)
- Ordinal encoding (Q-table)
- One-hot encoding (Double Q-learning)
- Ordinal encoding (Double Q-learning)
- One-hot encoding (Double Q-learning, Q-table)
- Ordinal encoding (Double Q-learning, Q-table)
- Modest solution = 75

X-axis: Number of episodes

Y-axis: Average reward