Consensus (N = 2) MDP Q-value, property steps_max, shallow 3-layered network

Average reward (y-axis, logarithmic scale from 10^{-1} to 10^4)

Number of episodes (x-axis, from 0 to 500)

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