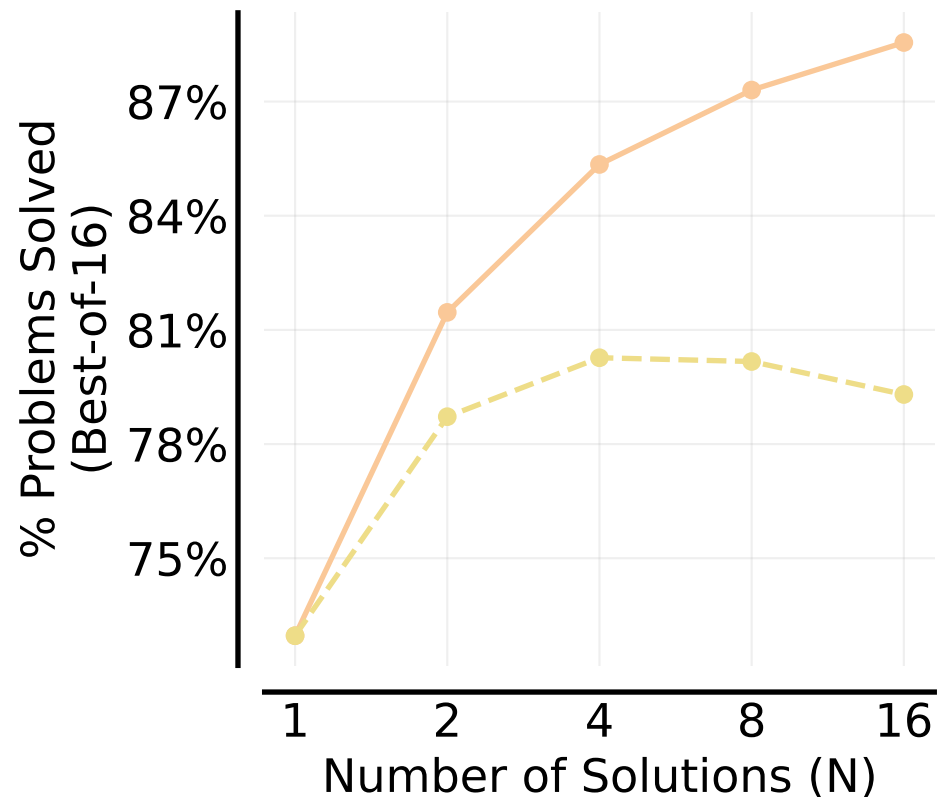


GSM8K: Ablation on DPO Reward



- DPO reward (without reference policy)
- -●- - DPO reward (with reference policy)