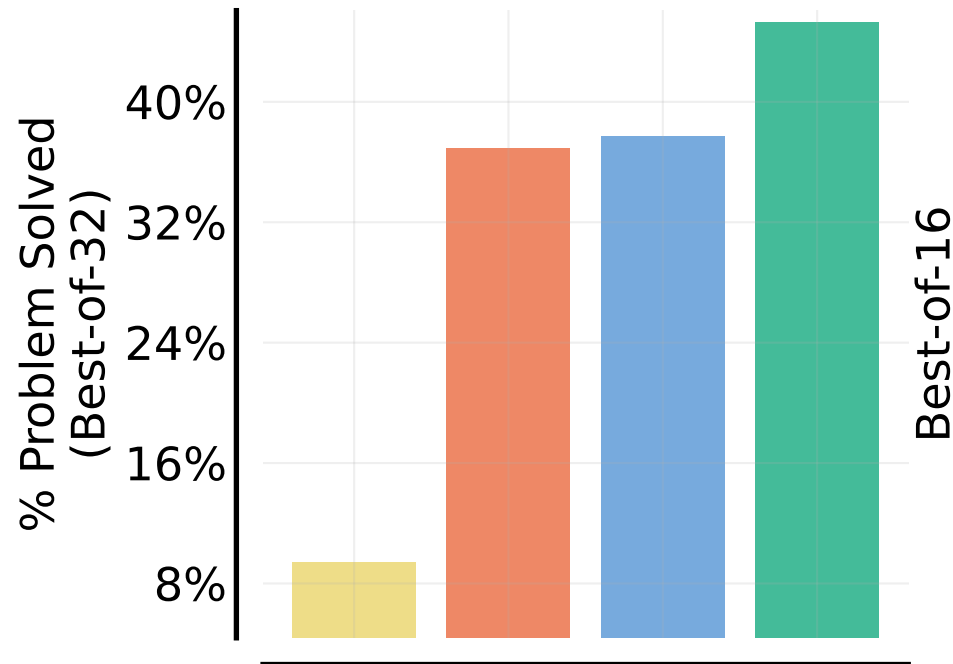




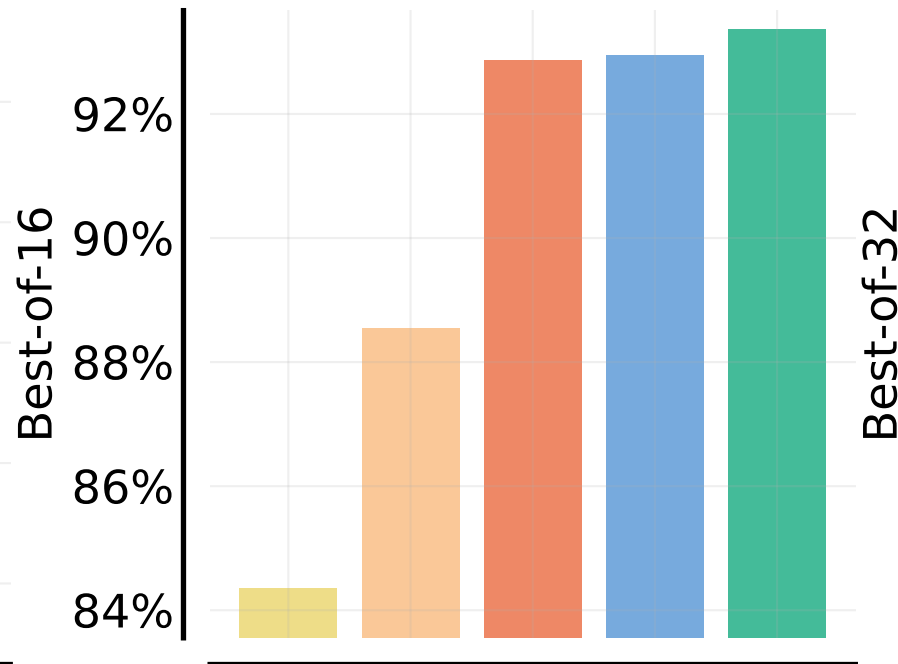
Algorithmic Reasoning (2 tasks)

$\Delta$ : 5.0%  $\rightarrow$  45.3%



Grade-School Math (GSM8K)

$\Delta$ : 73.0%  $\rightarrow$  93.4%



Transfer to MATH (GSM-Verifiers)

$\Delta$ : 28.0%  $\rightarrow$  44.6%

