Dr. Strategy: Model-Based Generalist Agents with Strategic Dreaming

Hany Hamed[1,*]   Subin Kim[1,*]   Dongyeong Kim[1]   Jaesik Yoon[1,2]   Sungjin Ahn[1]

[1]KAIST,   [2]SAP   * denotes equal contribution

Generalist Agents

A crucial capability of generalist agents, such as humans, is to explore environments and acquire the abilities needed to achieve various goals, continuously and in an open-ended way.
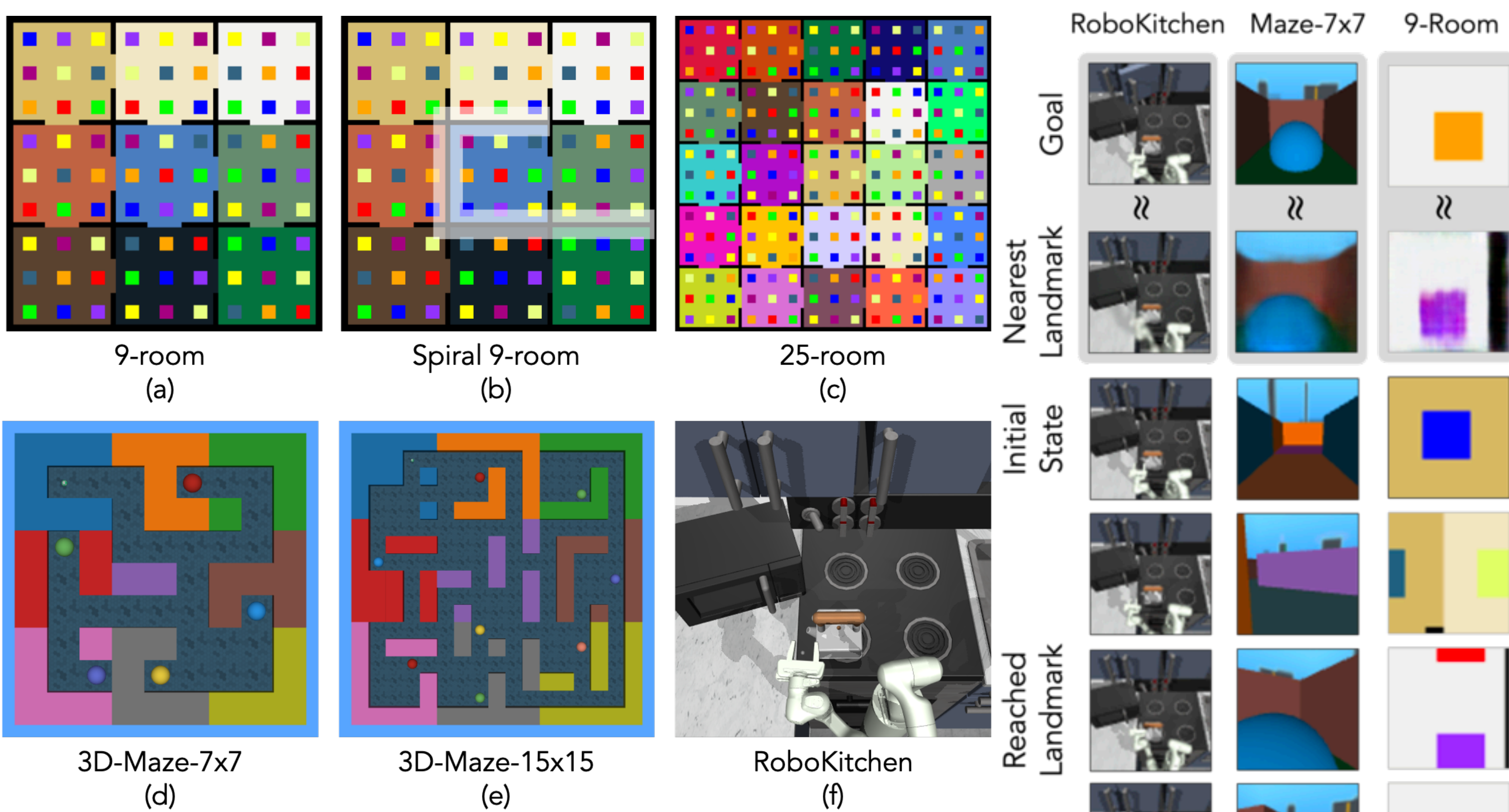
We raise the question of whether and how a model-based generalist RL agent can "dream better" in a more structured and strategic way.

Real world | Divide-and-Conquer | Point-to-Point

— By Train   ···· Forgotten path
— By Taxi   ○ Goal
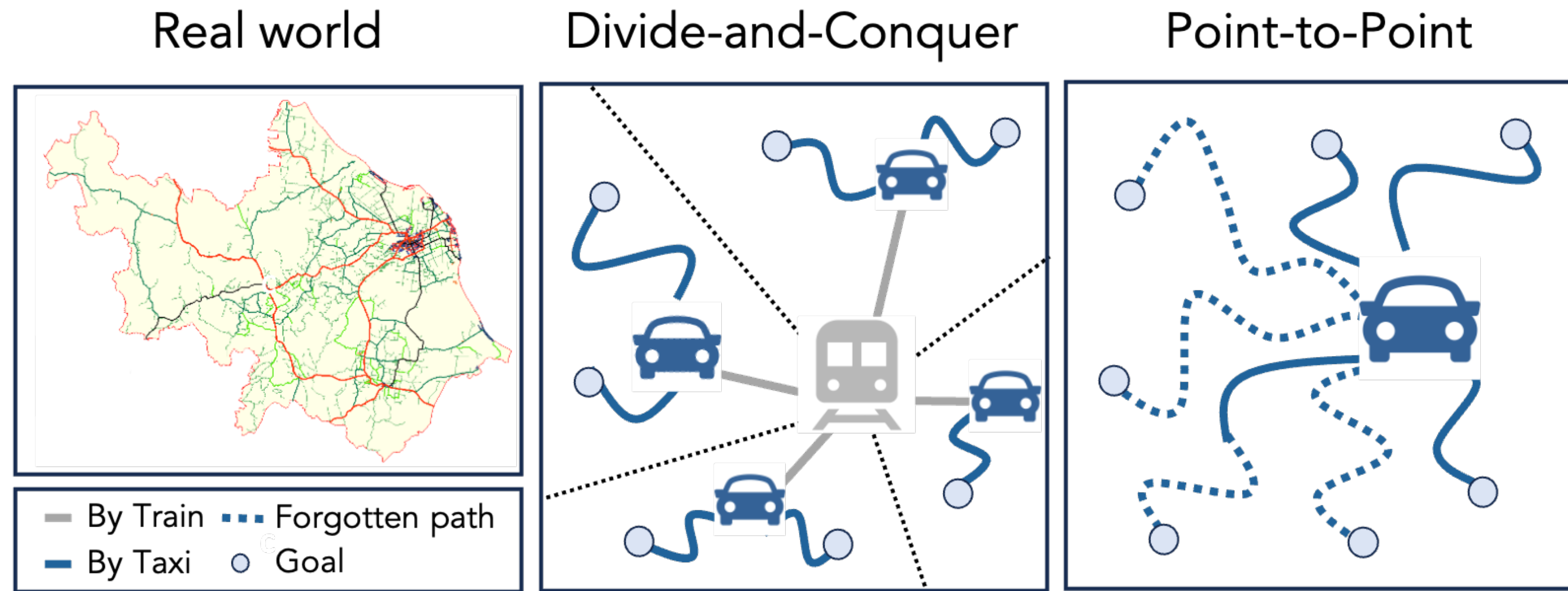
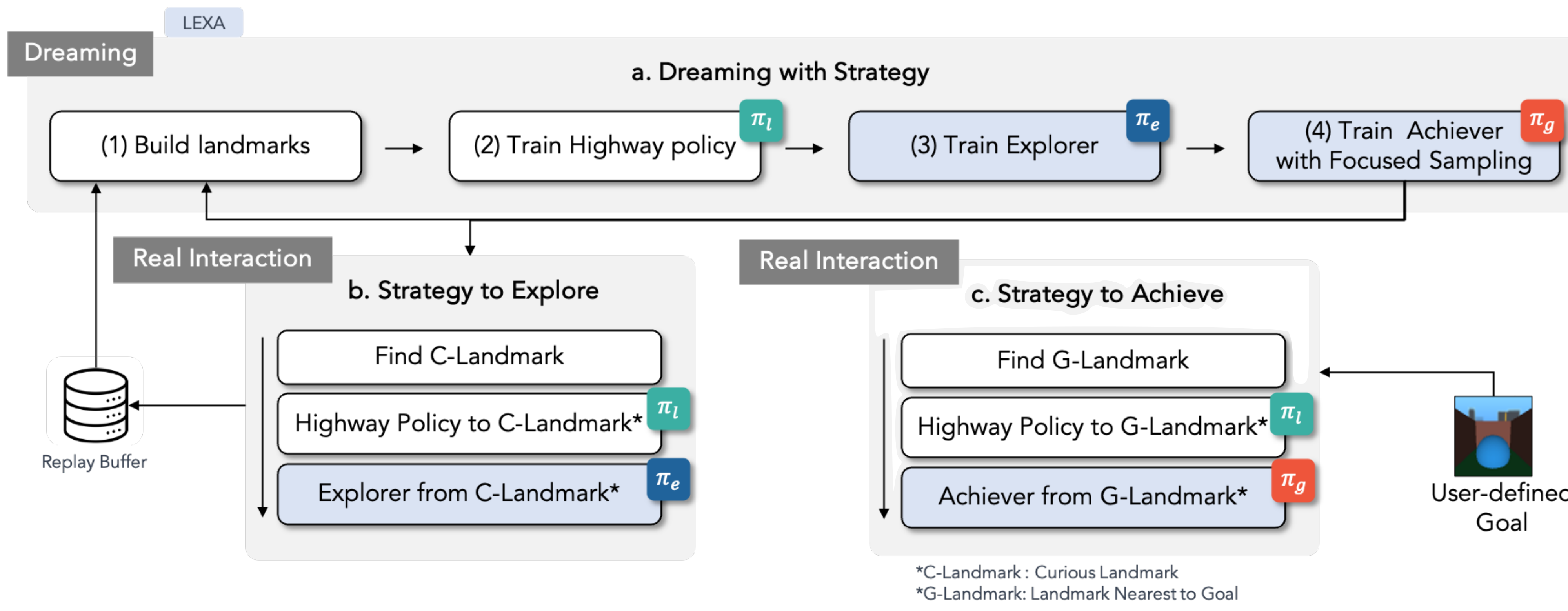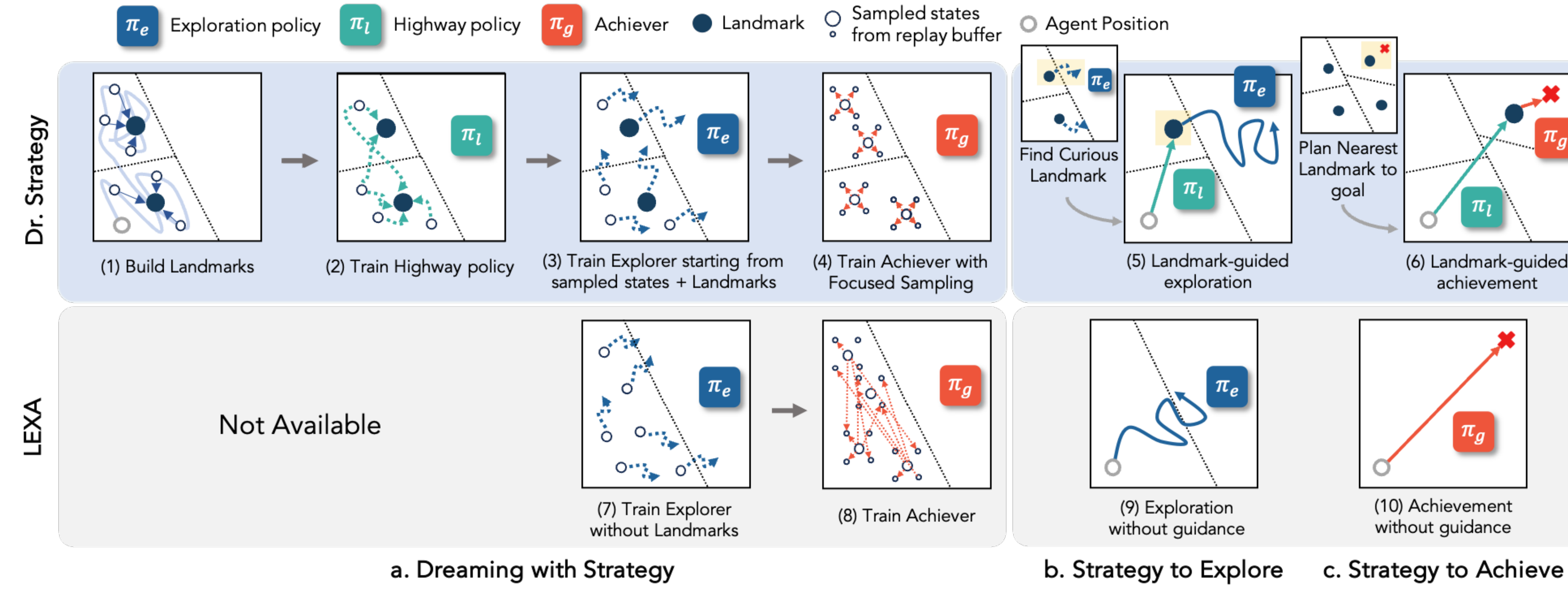**We propose a strategic model-based generalist agent, Dr. Strategy (short for "Dream Strategy")**

Our key idea is that a divide-and-conquer approach leveraging the structure of latent landmarks can enhance the efficiency of dreaming in MBRL and promote better exploration and achievement quality of a generalist agent.

Environments

9-room (a)   Spiral 9-room (b)   25-room (c)
3D-Maze-7x7 (d)   3D-Maze-15x15 (e)   RoboKitchen (f)
RoboKitchen   Maze-7x7   9-Room
Goal / Nearest Landmark / Initial State / Reached Landmark / Reached Goal — Time

Three types of visual-based environments:
(a-c) 2D POMDP navigation (partial view)
(d-e) 3D-Maze Navigation (ego-centric)
(f) a Franka Panda arm manipulation (third-person view)

Dr. Strategy

π_e Exploration policy π_l Highway policy π_g Achiever   ● Landmark   ○ Sampled states from replay buffer   ○ Agent Position

(1) Build Landmarks   (2) Train Highway policy   (3) Train Explorer starting from sampled states + Landmarks   (4) Train Achiever with Focused Sampling   (5) Landmark-guided exploration   (6) Landmark-guided achievement

Find Curious Landmark / Plan Nearest Landmark to goal

Not Available   (7) Train Explorer without Landmarks   (8) Train Achiever   (9) Exploration without guidance   (10) Achievement without guidance

a. Dreaming with Strategy   b. Strategy to Explore   c. Strategy to Achieve

a. Dreaming with Strategy
(1) Build landmarks → (2) Train Highway policy π_l → (3) Train Explorer π_e → (4) Train Achiever with Focused Sampling π_g

Dreaming / Real Interaction / Replay Buffer

b. Strategy to Explore
Find C-Landmark
Highway Policy to C-Landmark* π_l
Explorer from C-Landmark* π_e

c. Strategy to Achieve
Find G-Landmark
Highway Policy to G-Landmark* π_l
Achiever from G-Landmark* π_g
User-defined Goal

*C-Landmark : Curious Landmark
*G-Landmark: Landmark Nearest to Goal

Dreaming with Strategy

• Train the world model (DreamerV2)

Dynamics : $\hat{s}_t \sim \mathrm{dyn}_\theta(s_{t-1}, a_{t-1})$ Encoder : $e_t = \mathrm{enc}_\theta(x_t)$

Representation : $s_t \sim \mathrm{repr}_\theta(s_{t-1}, a_{t-1}, x_t)$ Decoder : $\hat{x}_t \sim \mathrm{dec}_\theta(s_t)$

• Learn the latent landmarks from the replay buffer using VQ-VAE

• Train three policies in imagination (Strategic Dreaming)

1) Highway policy: $r_l(s_t, l) = -\|\mathrm{dec}_\phi(l) - s_t\|_2^2 + \sum_{i=1}^{K} \log\|s_t - s_i^{\text{K-NN}}\|_2$

2) Explorer: $r_e(s_t) \propto Var(\{f(e_{t+1} \mid s_t, a_t, \theta_k) \mid k \in [1,\dots,N]\})$

3) Achiever: $r_g(\hat{e}_t, e_g) = -\mathrm{tdp}(\hat{e}_t, e_g)$

• Calculate the landmarks' curiosity: $C_i = \mathbb{E}_{\tau_i}[r_e(s_t^{(i)})], \quad \tau_i = \{s_0^{(i)}, s_1^{(i)}, \dots, s_H^{(i)}\}$
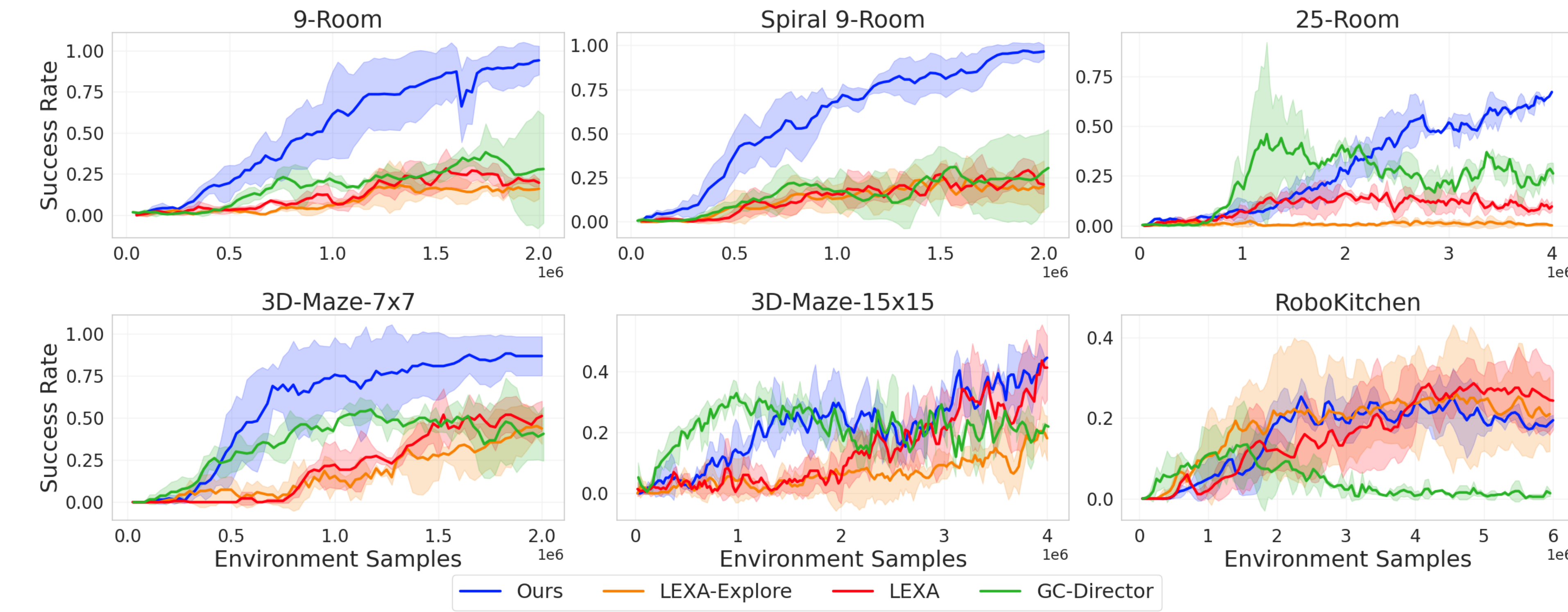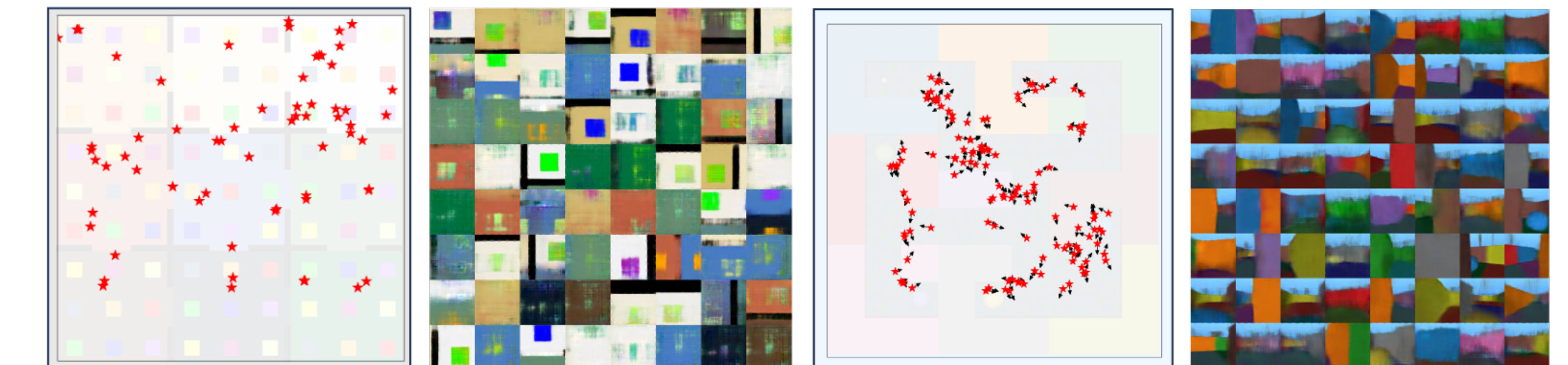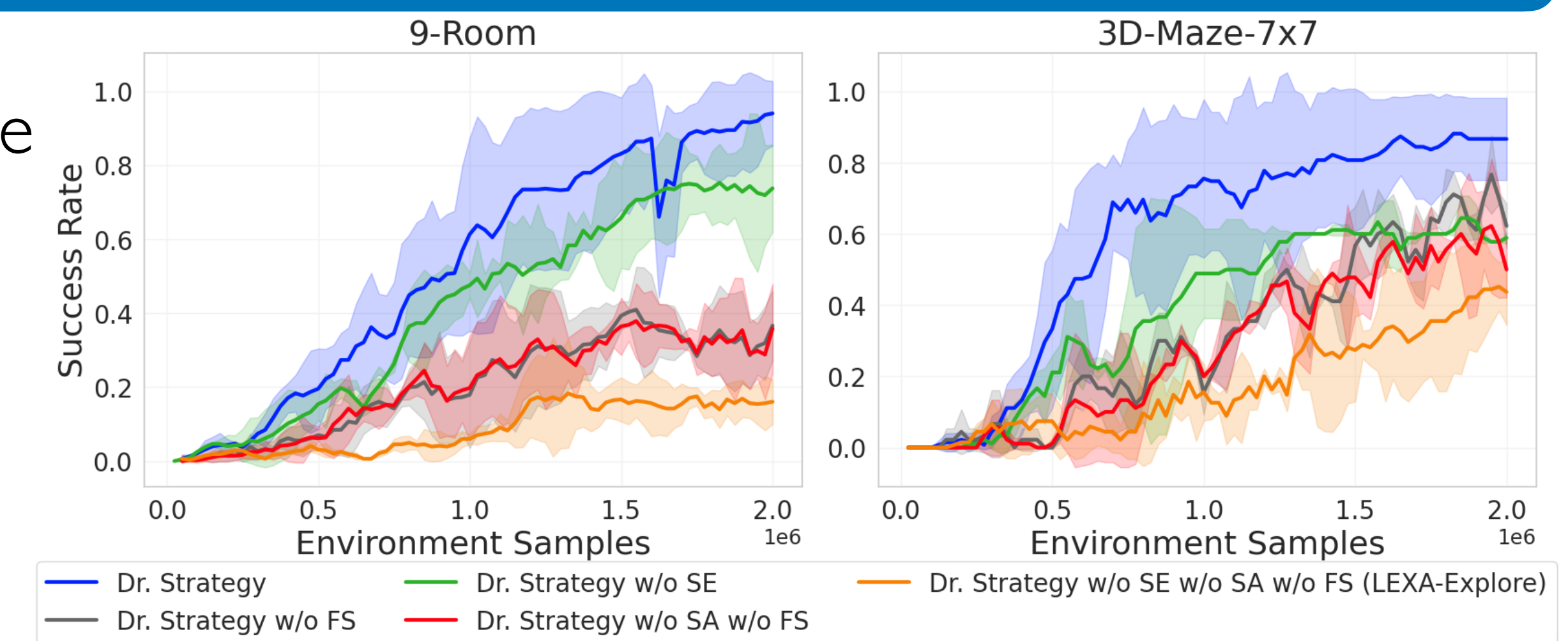
Results

9-Room | Spiral 9-Room | 25-Room
3D-Maze-7x7 | 3D-Maze-15x15 | RoboKitchen

Ours   LEXA-Explore   LEXA   GC-Director

Table 1. Final Success Rate

Method	9-Room	Spiral 9-Room	25-Room	Maze-7x7	Maze-25x25	RoboKitchen
LEXA	19.75%	21.19%	9.62%	51.11%	41.20%	**24.30%**
LEXA-Explore	16.04%	20.16%	0.14%	43.70%	18.05%	20.07%
GC-Director	28.08%	30.45%	27.11%	40.55%	21.87%	1.45%
Dr. Strategy (Ours)	**94.03%**	**96.50%**	**67.11%**	**86.66%**	**44.44%**	19.44%

Ablation Studies

We investigate the influence of three components of strategic dreaming:

Strategy to Explore (SE),
Strategy to Achieve (SA),
Focused Sampling (FS)

9-Room | 3D-Maze-7x7

Dr. Strategy   Dr. Strategy w/o SE   Dr. Strategy w/o SE w/o SA w/o FS (LEXA-Explore)
Dr. Strategy w/o FS   Dr. Strategy w/o SA w/o FS

Reconstructed Landmarks for 9-Room and 3D-Maze 7x7

Conclusion

• We propose Dr. Strategy, a novel model-based task generalist agent.

• Dr. Strategy outperformed prior pixel-based MBRL agents in various visually complex and partially observable navigation tasks, while also showing comparative results in robot manipulation tasks.