

Navigation Without a Finish Line — Infinite-Horizon Belief Maintenance for Robot Navigation

Arjen Singh

Jessie Chan

I. INTRODUCTION

We study persistent robot navigation under partial observability as an infinite-horizon belief-maintenance problem. Rather than reaching a single goal and terminating, the robot repeatedly receives new goals while using noisy observations and stochastic actions. The planner therefore must trade off immediate progress toward the current goal against maintaining a concentrated and reliable belief over the robot’s location. We formulate this setting as a POMDP and compare a myopic planner, which prioritizes task progress, with a belief-aware planner that adds an entropy penalty to encourage localization quality.

II. RELATED WORKS

Simmons and Koenig present a probabilistic robot navigation system that maintains a belief distribution over possible robot locations and updates it using sensor observations, actuator uncertainty, and map information [1]. Cassandra, Kaelbling, and Kurien similarly model mobile-robot navigation using discrete Bayesian decision models [2], while *Probabilistic Robotics* systematizes localization, mapping, and planning under uncertainty as core problems in autonomous systems [3]. More generally, Kaelbling, Littman, and Cassandra show that belief states are sufficient statistics for decision making in partially observable stochastic domains [4].

Because exact POMDP planning is computationally expensive, approximate solvers such as PBVI, SARSOP, POMCP, and DESPOT restrict computation to selected belief points, reachable belief spaces, Monte Carlo search, or sparse scenario trees [5], [6], [?], [7], [8]. These methods motivate our original belief-space planning formulation, although our final implementation uses a simpler map-based dynamic-programming value table combined with online belief updates.

Our work is most directly related to entropy-regularized and information-aware planning. Molloy and Nair study entropy-regularized POMDPs [9], and related work uses conditional entropy for active trajectory estimation [10]. These ideas motivate our belief-aware planner, which augments goal-directed value with an expected posterior entropy term.

III. PROBLEM FORMULATION

We formulate persistent robot navigation under partial observability as a POMDP $(\mathcal{S}, \mathcal{A}, \mathcal{O}, T, Z, r)$, following standard belief-space planning models [4], [3]. The robot does

not observe its true state $s_t \in \mathcal{S}$ directly, but maintains a belief $b_t \in \Delta(\mathcal{S})$ updated by

$$b_{t+1}(s') = \eta Z(o_{t+1} \mid s', a_t) \sum_{s \in \mathcal{S}} T(s' \mid s, a_t) b_t(s),$$

where η normalizes the distribution. Unlike a single-goal navigation task, our setting repeatedly samples new goals, so the robot must both reach the current goal and preserve localization quality for future tasks. We evaluate policies using long-run average reward,

$$J(\pi) = \liminf_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_\pi \left[\sum_{t=0}^{T-1} r(b_t, a_t) \right],$$

where

$$r(b_t, a_t) = R_g \mathbf{1}\{s_t \in G_t\} - \lambda_c \mathbf{1}\{s_t \in C\} - \lambda_H H(b_t) - \lambda_a c(a_t).$$

This reward explicitly values goal completion, collision avoidance, low belief entropy, and efficient motion.

The information-theoretic components within this entropy-regularized objective are defined as follows:

- **Belief Entropy $H(b_t)$:** This term quantifies the robot’s state uncertainty. Minimizing this entropy encourages a concentrated, “sharp” belief over possible states, but does not by itself guarantee that the belief is centered on the true state.
- **Belief Awareness Coefficient λ_H :** This parameter acts as the weight for belief regularization. Higher values of λ_H force the robot to actively prioritize localization quality, often driving the policy to take deliberate detours toward distinctive environmental landmarks to reduce uncertainty before proceeding to the goal.

IV. PROPOSED METHOD

A. Offline Component

The offline component approximates long-term navigation value before execution. In the original belief-space formulation, one natural approach would be to sample representative reachable beliefs from simulated rollouts and approximate a belief-value function $V(b)$ using point-based value iteration or a related sampled-belief dynamic programming method, following the motivation of PBVI and reachable-belief-space methods [5], [6]. This approach is attractive because exact POMDP planning is intractable in large domains. Restricting computation to sampled reachable beliefs would amortize expensive planning and focus computation on beliefs the robot is likely to encounter during execution.

In our final implementation, however, we use a simpler offline dynamic-programming backbone over physical grid states rather than solving a full sampled-belief PBVI problem. For each active goal g , the simulator computes a goal-conditioned state-value table $V_g(s)$ over the known map. This table estimates the utility of occupying each grid state while moving toward the current goal and avoiding obstacles. The full-observable planner uses this table directly because it has access to the true state. The partially observable planners use the same offline value table through the belief distribution by evaluating the expected navigation value,

$$\mathbb{E}_{s \sim b_t}[V_g(s)] = \sum_{s \in \mathcal{S}} b_t(s) V_g(s).$$

Here, $b_t(s)$ is the belief probability that the robot is in state s at time t , and $V_g(s)$ is the precomputed value of state s for the current goal g . The sum therefore computes the expected goal-directed value under the robot’s current localization uncertainty.

Thus, the key experimental comparison is not between different offline value solvers, but between using the same map-based dynamic-programming backbone with and without belief-maintenance regularization.

B. Online Component

The online component performs real-time belief updates and action selection during execution. At each timestep, the robot receives a noisy observation and updates its belief b_t over possible grid states using a Bayes filter. It then selects an action using the offline goal-conditioned value table described above. Since the robot does not know the true state under partial observability, the partially observable planners evaluate actions by their expected value under the current belief rather than by the value of a single known state. The myopic planner uses this expected navigation value directly, while the belief-aware planner augments the action score with an expected posterior entropy term so that actions are evaluated both by goal progress and by their effect on localization uncertainty.

This online procedure is a lightweight belief-state decision rule rather than a full multi-step online POMDP search. The planner uses the current belief, a predicted next belief, and a precomputed map-based value function to choose an action at each step. This makes the simulation computationally tractable while still testing the central experimental question: whether adding belief-maintenance regularization improves persistent navigation compared with a myopic belief-state baseline.

To assess how effectively this belief-state action-selection architecture balances task progress, tracking quality, and empirical safety, we separate raw task performance from the entropy-aware planning objective during evaluation. Rather than claiming that objective reward tuning guarantees safe execution, we report goal completion, task reward, objective reward, belief entropy, probability assigned to the true state, maximum a posteriori (MAP) error, collisions, invalid actions, and steps per goal. These metrics are evaluated

across three diagnostic environments: Symmetric-Aliasing layouts designed to expose multimodal tracking ambiguity, Asymmetric-Control layouts where goal-directed behavior should dominate, and Bottleneck-Safety layouts with narrow passages that test whether localization improvements translate into safer motion under uncertainty.

V. EXPERIMENTAL RESULTS

A. Simulation Setup and Experimental Design

Our simulation platform evaluates persistent robot navigation under partial observability in finite grid-world maps with obstacles, landmarks, stochastic motion, noisy observations, and sequentially sampled goals. The robot must repeatedly reach goals while maintaining a useful belief distribution over its possible locations. Each diagnostic run uses a 200-step horizon, transition noise $p_{\text{trans}} = 0.25$, and observation noise $\sigma_{\text{obs}} = 0.5$. The transition noise parameter controls stochastic deviations from the commanded action, while the observation noise parameter controls the uncertainty in the robot’s landmark-based sensor measurements.

The core hypothesis is that adding a nonzero belief-awareness coefficient ($\lambda_H > 0$) can improve long-run belief quality in ambiguous environments compared with a myopic goal-directed baseline ($\lambda_H = 0$). However, the experiments also test whether improved belief quality actually translates into better task throughput and safer motion, since entropy reduction alone does not formally guarantee correct localization or obstacle avoidance.

To evaluate this, we compare planners across three map layouts designed to isolate different aspects of belief-space planning:

- 1) **Symmetric-Aliasing Environment:** This layout consists of two mirrored rooms around a central landmark. It tests whether the planner can resolve multimodal tracking ambiguity when different physical states produce similar observations.
- 2) **Asymmetric-Control Environment:** This layout introduces additional disambiguating structure and serves as a lower-ambiguity control setting where goal-directed navigation should dominate.
- 3) **Bottleneck-Safety Environment:** This layout contains narrow passages and constrained corridors, testing whether localization improvements translate into safer motion near obstacles.

B. Baselines and Quantitative Performance

We evaluate two partially observable planner configurations: a **Myopic Planner** ($\lambda_H = 0$), which prioritizes expected goal-directed value under the current belief, and a **Belief-Aware Planner** ($\lambda_H = 0.5$), which adds an expected posterior entropy term to the planning objective. We also include a **Full Observable** planner as an upper-reference baseline that has access to the true robot state. To decouple raw navigation performance from belief quality, we report two groups of metrics: task performance and efficiency in Table I, and belief-inference quality and safety in Table II.

The 200-step diagnostic results show that belief awareness does not uniformly dominate the myopic baseline. Instead, its effect depends strongly on the structure of the environment. In the Asymmetric-Control map, where the added landmark reduces perceptual aliasing, the belief-aware planner slightly improves localization quality while preserving the same goal throughput as the myopic planner. In the Symmetric-Aliasing map, however, the belief-aware planner reduces mean entropy but worsens MAP error and completes fewer goals, suggesting that entropy minimization can produce a concentrated but incorrect belief. In the Bottleneck-Safety map, the belief-aware planner improves goal completion and MAP error, but also increases collisions and invalid attempts. Thus, the experiments reveal a tradeoff among information gathering, task progress, and safety rather than a single planner being uniformly best.

C. Analysis of the Symmetric-Aliasing Environment

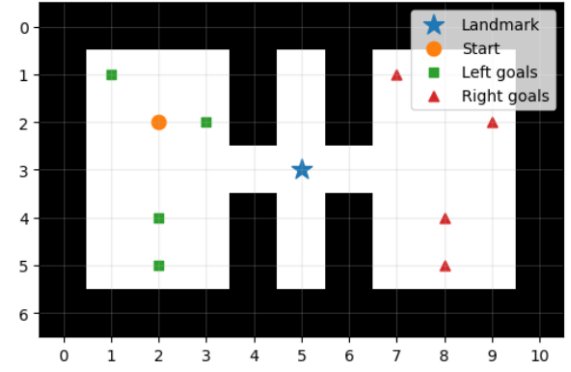
In the Symmetric-Aliasing map layout, which consists of two mirrored rooms centered around a single landmark as shown in Figure 1, multiple physical states can generate similar observations. This environment directly tests whether entropy regularization can resolve perceptual aliasing. The fully observable planner reaches 17 goals with no collisions, establishing the upper-reference performance when localization uncertainty is removed. By contrast, the myopic planner reaches 5 goals, while the belief-aware planner reaches only 2 goals, indicating that the entropy-aware objective reduces task throughput in this highly symmetric map.

The belief-aware planner does reduce mean entropy slightly, from 0.8490 for the myopic planner to 0.7959, and assigns slightly higher average probability to the true state, increasing $P(\text{true})$ from 0.4617 to 0.4729. The rolling entropy plot in Figure 1b supports this trend: after the initial drop, the belief-aware curve settles near a smoother and slightly lower entropy level, while the myopic planner fluctuates more sharply as it alternates between periods of confidence and ambiguity. However, the MAP error shows that this smoother entropy profile is not necessarily more accurate. The myopic planner has a MAP error of 0.580, while the belief-aware planner has a much larger MAP error of 1.825. This suggests that the belief-aware planner becomes more certain while still being wrong, likely collapsing its belief onto an incorrect aliased mode. Therefore, the Symmetric-Aliasing experiment is best interpreted as a failure case for entropy regularization alone: minimizing uncertainty does not guarantee that the concentrated belief is centered on the true robot state.

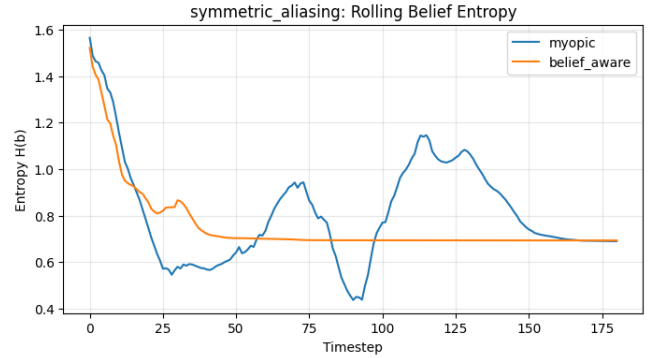
D. Stress-Testing and Safety Analysis

To further stress-test the robustness of the belief-maintenance framework, we evaluate performance across the Asymmetric-Control and Bottleneck-Safety environments. The simulation setup is intentionally diagnostic rather than purely benchmark-oriented: each map isolates a different failure mode of persistent navigation under partial observability. The Symmetric-Aliasing map tests perceptual aliasing, the

symmetric_aliasing: Two mirrored rooms with one central landmark.



(a) Symmetric-Aliasing map layout.



(b) Symmetric-Aliasing rolling belief entropy.

Fig. 1: Structural map layout and corresponding internal filter profile over 180 timesteps for the Symmetric-Aliasing condition.

Asymmetric-Control map tests low-ambiguity goal pursuit, and the Bottleneck-Safety map tests whether localization errors become physically costly in constrained passages. This allows us to evaluate not only whether the robot completes goals, but also whether the belief state remains accurate, whether entropy is a reliable proxy for correctness, and whether improved localization translates into safer motion.

In the Asymmetric-Control layout, shown in Figure 2a, an additional disambiguating landmark reduces the symmetry present in the mirrored-room map. As a result, both partially observable planners perform close to the fully observable reference: the myopic and belief-aware planners each complete 16 goals, compared with 17 for the fully observable planner. The belief-aware planner slightly improves belief quality, reducing mean entropy from 0.4756 to 0.4512, increasing $P(\text{true})$ from 0.7335 to 0.7437, and reducing MAP error from 0.215 to 0.180. The rolling entropy plot in Figure 2b supports this interpretation: the belief-aware planner maintains slightly lower uncertainty over much of the trajectory, although the myopic planner briefly achieves lower entropy near the middle of the run. Thus, in this less ambiguous environment, entropy regularization acts as a mild stabilizer rather than causing the throughput loss observed in the Symmetric-Aliasing case.

The Bottleneck-Safety layout, shown in Figure 3a, creates

TABLE I: Task performance and efficiency over the 200-step diagnostic simulation. Higher goals and rewards are better ($\uparrow$); fewer steps per goal is better ($\downarrow$).

Environment	Planner	Goals $\uparrow$	Task R/Step $\uparrow$	Obj. R/Step $\uparrow$	Steps/Goal $\downarrow$
Symmetric-Aliasing	Full observable	17	8.0000	8.0000	11.7647
	Myopic	5	2.2900	1.8655	40.0000
	Belief-aware	2	0.8825	0.4846	100.0000
Asymmetric-Control	Full observable	17	8.0000	8.0000	11.7647
	Myopic	16	7.4200	7.1822	12.5000
	Belief-aware	16	7.4150	7.1894	12.5000
Bottleneck-Safety	Full observable	17	8.0000	8.0000	11.7647
	Myopic	4	1.7675	1.3483	50.0000
	Belief-aware	6	2.6425	2.1029	33.3333

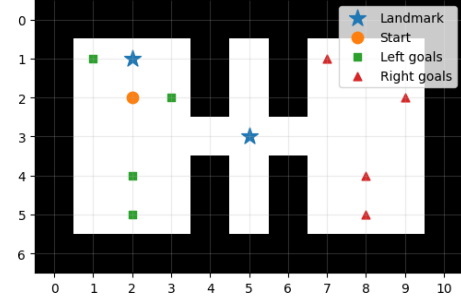
TABLE II: Belief-inference quality and safety over the 200-step diagnostic simulation. Lower entropy, MAP error, collisions, and invalid attempts are better ($\downarrow$); higher $P(\text{true})$ is better ($\uparrow$).

Environment	Planner	Mean H $\downarrow$	$P(\text{True})$ $\uparrow$	MAP Err. $\downarrow$	Coll. $\downarrow$	Invalid $\downarrow$
Symmetric-Aliasing	Full observable	0.8078	0.5540	0.455	0	0
	Myopic	0.8490	0.4617	0.580	4	6
	Belief-aware	0.7959	0.4729	1.825	4	5
Asymmetric-Control	Full observable	0.4742	0.7390	0.195	0	0
	Myopic	0.4756	0.7335	0.215	9	11
	Belief-aware	0.4512	0.7437	0.180	9	10
Bottleneck-Safety	Full observable	0.8853	0.5594	0.355	0	0
	Myopic	0.8384	0.4617	1.440	9	11
	Belief-aware	1.0792	0.4347	0.290	12	15

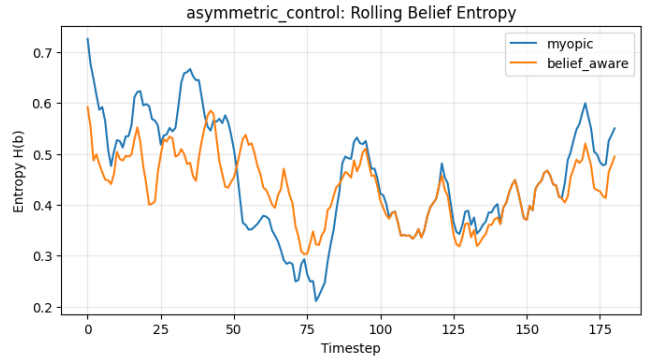
a different tradeoff because narrow passages make localization errors physically costly. In this environment, the belief-aware planner improves task throughput and localization accuracy: it completes 6 goals compared with 4 for the myopic planner, and reduces MAP error from 1.440 to 0.290. However, it does not improve the empirical safety metrics. Actual collisions increase from 9 to 12, and invalid attempts increase from 11 to 15. The rolling entropy plot in Figure 3b further shows that the belief-aware planner does not simply reduce uncertainty throughout the run; instead, its entropy remains higher than the myopic planner for much of the trajectory, with pronounced spikes near the middle of the run. This suggests that the belief-aware planner may enter more ambiguous or constrained regions while seeking informative observations. Although this improves average state estimation accuracy, it also increases contact with obstacles.

These stress tests clarify the main limitation of the current planner. In informative open layouts, entropy regularization can stabilize the belief without hurting goal completion. In symmetric environments, it may collapse the belief onto the wrong mode. In narrow bottlenecks, it may improve localization while worsening collision behavior. Therefore, entropy shaping should be treated as a belief-quality heuristic rather than a formal safety mechanism, motivating future extensions with explicit collision-aware costs, chance constraints, or safety filters.

asymmetric_control: The same geometry with an extra disambiguating landmark.



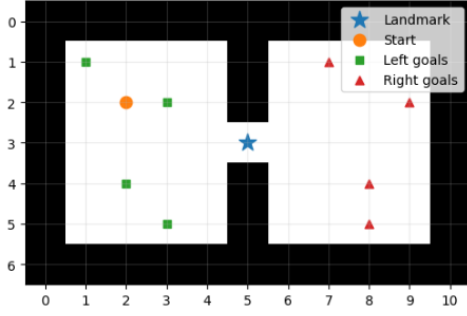
(a) Asymmetric-Control map layout.



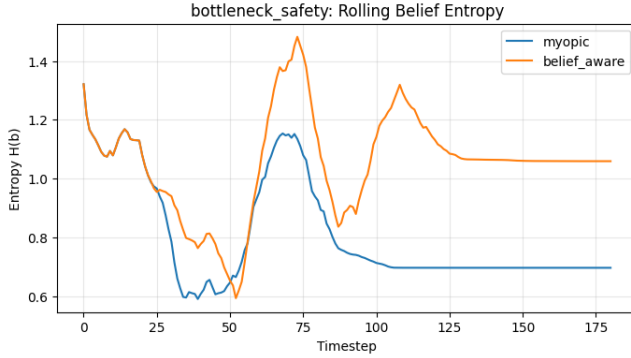
(b) Asymmetric-Control rolling belief entropy.

Fig. 2: Structural map layout and corresponding internal filter profile over 180 timesteps for the Asymmetric-Control condition.

bottleneck_safety: Two rooms connected through a narrow informative passage.



(a) Bottleneck-Safety map layout.



(b) Bottleneck-Safety rolling belief entropy.

Fig. 3: Structural map layout and corresponding internal filter profile over 180 timesteps for the Bottleneck-Safety condition.

VI. CONCLUSION & FUTURE WORK

This work studies persistent robot navigation under partial observability as an infinite-horizon belief-maintenance problem. The updated simulations show that entropy-regularized planning creates a real but delicate tradeoff among information gathering, goal completion, and safety, consistent with prior work showing that information-theoretic terms can shape behavior in partially observed decision problems [9], [10], [11]. In the Asymmetric-Control environment, belief awareness improves localization quality while preserving the same goal throughput as the myopic planner. However, in the Symmetric-Aliasing environment, the belief-aware planner achieves lower mean entropy but worse MAP error and fewer goals, showing that a lower-entropy belief is not necessarily a more accurate belief. In the Bottleneck-Safety environment, belief awareness improves goal completion and MAP error, but increases collisions and invalid attempts.

The main conclusion is therefore not that entropy regularization is universally superior. Rather, entropy regularization is useful only when the environment and planner can convert information-seeking behavior into correct localization and safe motion. Belief entropy is a helpful diagnostic metric, but it is insufficient by itself because it does not distinguish between useful certainty and overconfident belief in the wrong state. Similarly, better localization does not automatically guarantee collision avoidance, especially in narrow passages

where small errors can have large safety consequences.

Future work should address these limitations at the planning level rather than relying only on reward tuning. First, the online planner should be extended with a deeper look at the belief-space through MCTS or POMCP so that information-gathering actions are evaluated based on their long-term navigation value rather than the immediate reduction of entropy alone. Second, the safety objective should be made explicit through stronger collision penalties, chance constraints, control-barrier-style constraints, or reachability-based methods, since empirical entropy reduction does not provide a formal safety guarantee [12], [13], [14]. Third, the evaluation should be expanded from single-seed diagnostic runs to multi-seed averages across transition noise, observation noise, entropy weight, collision weight, and action-cost weight, following the broader evaluation philosophy used in approximate POMDP solvers such as PBVI, SARSOP, POMCP, and DESPOT [5], [6], [15], [7]. Finally, future versions should report calibration-style belief metrics in addition to entropy, so that the planner can be evaluated not only on how certain it is, but also on whether that certainty is correct.

REFERENCES

- [1] R. Simmons and S. Koenig, "Probabilistic robot navigation in partially observable environments," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 1995, pp. 1080–1087.
- [2] A. R. Cassandra, L. P. Kaelbling, and J. A. Kurien, "Acting under uncertainty: Discrete bayesian models for mobile-robot navigation," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1996.
- [3] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. MIT Press, 2005.
- [4] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial Intelligence*, vol. 101, no. 1–2, pp. 99–134, 1998.
- [5] J. Pineau, G. Gordon, and S. Thrun, "Point-based value iteration: An anytime algorithm for POMDPs," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2003, pp. 1025–1032.
- [6] H. Kurniawati, D. Hsu, and W. S. Lee, "SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces," in *Robotics: Science and Systems*, 2008.
- [7] A. Somani, N. Ye, D. Hsu, and W. S. Lee, "DESPOT: Online POMDP planning with regularization," in *Advances in Neural Information Processing Systems*, 2013.
- [8] N. Ye, A. Somani, D. Hsu, and W. S. Lee, "DESPOT: Online POMDP planning with regularization," *Journal of Artificial Intelligence Research*, vol. 58, pp. 231–266, 2017.
- [9] T. L. Molloy and G. N. Nair, "Entropy-regularized partially observed markov decision processes," *IEEE Transactions on Automatic Control*, vol. 69, no. 1, pp. 379–386, 2024.
- [10] —, "Active trajectory estimation for partially observed markov decision processes via conditional entropy," in *Proceedings of the European Control Conference (ECC)*, 2021, pp. 385–391.
- [11] T. Tanaka, H. Sandberg, and M. Skoglund, "Transfer-entropy-regularized markov decision processes," *IEEE Transactions on Automatic Control*, vol. 67, no. 4, pp. 1944–1951, 2022.
- [12] A. D. Ames, S. Coogan, M. Egerstedt, G. Notomista, K. Sreenath, and P. Tabuada, "Control barrier functions: Theory and applications," *European Journal of Control*, vol. 45, pp. 1–23, 2019.
- [13] I. M. Mitchell, A. M. Bayen, and C. J. Tomlin, "A time-dependent hamilton-jacobi formulation of reachable sets for continuous dynamic games," in *IEEE Transactions on Automatic Control*, vol. 50, no. 7, 2005, pp. 947–957.
- [14] L. Lindemann, M. Cleaveland, G. Shim, and G. J. Pappas, "Safe planning in dynamic environments using conformal prediction," *IEEE Robotics and Automation Letters*, vol. 8, no. 8, pp. 5116–5123, 2023.

- [15] D. Silver and J. Veness, “Monte-carlo planning in large pomdps,” in *Advances in Neural Information Processing Systems*, 2010.