RoCUS: Robot Controller Understanding via Sampling

Anonymous Author(s)
Affiliation
Address
email

Abstract: As robots are deployed in complex situations, engineers and end users must develop a holistic understanding of their behaviors, capabilities, and limitations. Some behaviors are directly optimized by an objective, but typically many are emergent. The behaviors which are most often directly optimized relate to task completion: success rate, completion time, or total energy consumption. Other behaviors like collision avoidance, trajectory smoothness, and motion legibility are equally important for safe and trustworthy deployment, but are less often explicitly optimized. Designing an objective which optimizes every aspect of robot behavior is hard, so we instead propose a method for systematic analysis of exhibited behaviors. We present RoCUS: a Bayesian sampling-based method to find situations that lead to robot trajectories which exhibit specified behaviors—for example, highly jerky motions. For a 2D navigation problem and a 7 degree-of-freedom (DoF) arm reaching problem, we use RoCUS to analyze and compare three controllers: a deep imitation learning (IL) or reinforcement learning (RL) model, a rapidly exploring random tree (RRT) planner, and a dynamical system (DS) formulation. We show how RoCUS can uncover insights to further our understanding about these controllers beyond task-completion aspects and lead to better controller design. Code and video are submitted as supplementary files.

Keywords: Debugging and Evaluation, Algorithmic Transparency

1 Introduction

In 2018, after a confluence of failures, an autonomous vehicle (AV) struck and killed a pedestrian for the first time. In the run-up to this fateful event, the responsible company had reportedly been trying to improve the AV "ride experience" by emphasizing non-critical behaviors—such as the smoothness of the ride [1]. This event reflects the long-standing challenge in robotics of designing an appropriate objective which considers both safety-critical and non-critical behaviors. When crafting an objective, it is virtually impossible to proactively account for all potential controller behaviors, and some priorities may even be in conflict with one another [2]. In practice, any given robot behaviors may be specified, unspecified, or even misspecified [3], so extensive testing and evaluation is a critical component of designing and assessing robot controllers—especially learned controllers.

A common form of testing focuses on finding extreme and edge cases of controller failure. For example, a tester might use this procedure to find that the AV swerves very badly when encountering a farm animal while traveling at 60mph. Finding such extreme and edge cases is well-studied within both traditional software testing paradigms [4] and more recent adversarial perturbation testing methods [5]. However, we argue that an equally, if not more, important form should focus on *representative* scenarios, which considers the likelihood of encountering these scenarios. For example, if this AV is going to be deployed exclusively in New York City, the above example is largely unhelpful: cars rarely travel at 60mph in the city, and are very unlikely to encounter farm animals. Instead, the tester may prefer to know that the car swerves—though not as substantively—at lower speeds when a pedestrian steps toward it. Finding representative scenarios is often overlooked, but is especially useful for robotics. This is the focus of this paper.

Explicit mathematical analysis of robot controllers is implausible given the high dimensionality of the configuration space and the potential black-box representation of a learned controller. With access to an environment simulator, though, a straightforward testing approach is to roll out the robotic controller on various environments (e.g. road conditions under different weather and con-

Submitted to the 5th Conference on Robot Learning (CoRL 2021). Do not distribute.

gestion, with or without farm animals or pedestrians, etc.), and analyze those rollouts that exhibit a specified behavior—like excessive swerving. However, with too few environments, we risk missing the condition(s) that triggers the target behavior most saliently. With too many environments, all the most salient rollouts would be close to the global maximum at the expense of diversity and coverage. For example, if a farm animal causes the most swerving, followed by a pedestrian and a dangling tree branch, using too few environments may only find the pedestrian and the tree branch while using too many would result in an exclusive focus on the farm animal. Neither case helps the human develop a correct mental model of the AV's behavior.

To address this, we introduce Robot Controller Understanding via Sampling (ROCUS), a method to enable systematic behavior inspection. Using a posterior sampling procedure, ROCUS finds scenarios that are both inherently likely and which elicit specified behaviors. Analyzing these scenarios and the resulting trajectories can help engineers, testers, and end users inform their understanding of robot behaviors, and iterate on algorithm development if undesirable behaviors are revealed.

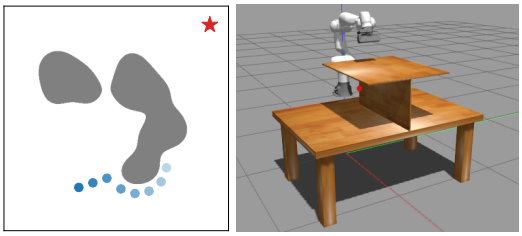

Figure 1: We analyze two problems using ROCUS. Left: 2D navigation tast, where the robot needs to reach the red star target while avoiding irregularly-shaped gray obstacles, e.g., by following the blue path. Right: 7DoF arm reaching task, where the robot needs to reach a given target (small red sphere on the left) which may be on either side under the T-shaped divider.

We use ROCUS to analyze three controllers on two common robotics tasks (Fig. 1). For a 2D navigation problem, we consider imitation learning (IL) [6], dynamical system (DS) [7], and rapidly-exploring random tree (RRT) [8]. For a 7DoF arm reaching problem, we consider reinforcement learning (RL) [9], as well as the same DS and RRT controllers. For each problem and controller, we specify several behaviors and visualize representative tasks and trajectories that elicit those behaviors. Through this analysis, we uncover insights that would be hard to derive analytically—insights which complement our mathematical understanding of the controllers. As such, ROCUS is a step towards the broader goal of building more accurate human mental models and enabling holistic evaluation of robot behaviors. Moreover, we include a case study on how to improve a controller based on new insights from ROCUS.

2 Related Work

Our work lies at the intersection of efforts to understand complex model behaviors and to benchmark robot performance. Methods to understand, interpret, and explain model behaviors are now commonplace in the machine learning community. [10] introduced Model Cards, a model analysis mechanism which breaks down model performance for data subsets. In natural language processing, [11] introduced a checklist for holistic evaluation of model capabilities and test case generation. [12] introduced BAYES-TREX, a Bayesian inference framework for sampling specified classifier behaviors. In robotics, [13] introduced a verification framework for assessing machine behavior by sampling parameter spaces to find temporal logic-satisfying behaviors. Other efforts aim to summarize robot policies, trading off factors like brevity, diversity, and completeness [14, 15]. All of these works have a shared underlying approach: treating the black box as immutable and performing downstream analyses of machine behavior [16]. ROCUS shares this approach, and, similar to [12], searches for instantiations which exhibit target behaviors to inform accurate human mental models.

While the need for benchmarking robot performance is often expressed [17, 18, 19], these efforts usually operate on distributions of trajectories or randomly selected trajectories, and the accompanying metrics are typically task-completion based without consideration of implicit performance factors. [20] put forth a recommendation of using *success weighted by path length* for navigation tasks—a task-completion metric. [21] and [22] introduced suites of metrics for comparing motion planning approaches, and [23] introduced a set of task and motion planning scenarios and metrics. Again, all of these proposed metrics are based solely on task completion. [24] introduced a set of performance measures for benchmarking reaching tasks; these benchmarks are either task-completion based or require a costly human motion ground truth. Our contribution is distinct in two ways. First, we propose to sample specific trajectories which communicate controller behaviors instead of reporting metrics averaged over distributions of trajectories. Second, we introduce metrics which draw on these prior works while also including essential alternative and typically emergent quality factors, like motion jerkiness and legibility [25].

3 RoCUS

RoCUS helps users understand robotic controllers via representative scenarios that exhibit various specified behaviors. It solves this by directly incorporating the distribution of scenarios—formally, *tasks*—into a Bayesian inference framework as shown in Fig. 2. A robotic problem is represented by a distribution $\pi(t)$ of individual tasks t. For example, a navigation problem may have $\pi(t)$ representing the distribution over target locations and obstacle configurations. Given a specific task t, the controller of interest induces a distribution $p(\tau|t)$ of possible trajectories τ. If both the controller and the transition dynamics are deterministic, $p(\tau|t)$ reduces to a δ-function at the induced trajectory τ. Stochasticity in either the controller (e.g., RRT) or the dynamics (e.g., uncertain outcome from an action) can result in τ being random. Finally, a behavior function $b(\tau, t)$ computes the behavior value of the trajectory—for example, the motion jerkiness. Some behaviors only depend on the trajectory and not the task, but we use $b(\tau, t)$ for consistency. Sec. 4 presents a list of behaviors.

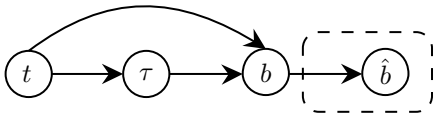

Figure 2: The graphical model for the inference problem of finding tasks t and trajectories τ which exhibit specific behaviors b. The dashed box indicates the relaxed formulation (Eq. 2).

The discussion on behavior in Sec. 1 is informal and implicitly combines two related but different concepts. The first concept is the behavior function $b(\tau, t)$ discussed above. The second is the specified target: for the swerving example, we are particularly interested in maximal behavior values. Thus, the target value can be thought of as $+\infty$. This inference problem uses the *maximal* mode of RoCUS. In other cases, we are also interested in tasks and trajectories that exhibit a specific finite behavior value. For example, we want to find road conditions that lead to a daily commute time of an hour, where the behavior is the travel time. This inference problem uses the *matching* mode.

Since matching mode is conceptually simpler, we present it first, followed by maximal mode. The sampling procedure is the same for both modes and presented last (Alg. 1).

3.1 Matching Mode

The exact objective is to find tasks and trajectories that exhibit user-specified behaviors b^*:

$$t, \tau \sim p(t, \tau | b = b^*) \propto p(b = b^* | t, \tau)\pi(\tau|t)\pi(t). \tag{1}$$

In most cases this posterior does not admit direct sampling, and an envelope distribution is not available for rejection sampling. Markov-Chain Monte-Carlo (MCMC) sampling does not work either: since the posterior is only non-zero on a very small or even measure-zero set, a Metropolis-Hastings (MH) sampler [26] can be stuck in the zero-density region.

Similar to [12], we relax the formulation as shown in Fig. 2:

$$\widehat{b}|b \sim \mathcal{N}(b, \sigma^2) \qquad t, \tau \sim p(t, \tau | \widehat{b} = b^*) \propto p(\widehat{b} = b^* | t, \tau) p(\tau|t)\pi(t). \tag{2}$$

This relaxed posterior is non-zero everywhere $\pi(t)$ is non-zero and provides useful guidance to an MH sampler. While σ is a hyper-parameter in [12], we instead assign σ such that

$$\int_{b^*-\sqrt{3}\sigma}^{b^*+\sqrt{3}\sigma} p(b)\,\mathrm{d}b = \alpha, \;\; \text{with} \;\; p(b) = \int_t \int_\tau p(\tau|t)\pi(t)\mathbb{1}_{b(\tau,t)=b}\,\mathrm{d}\tau\,\mathrm{d}t \tag{3}$$

being the marginal distribution of $b(\tau, t)$. Since we do not have a closed-form solution to $p(b)$, we can approximate it with its empirical count distribution, $\widehat{p}(b) = \sum_{i=1}^{N_B} \mathbb{1}_{b=b_i}/N_B$ for $\{b_1, ..., b_{N_B}\}$ drawn from the marginal distribution.

This formulation has two desirable properties. First, it is scale-invariant with respect to $b(\tau, t)$. For example, the posterior distribution is the same regardless of whether the trajectory length is measured in meters or centimeters. Second, the hyper-parameter $\alpha \in [0, 1]$ is intuitively interpreted as the approximate "volume" of posterior samples t, τ where $\widehat{b} = b^*$ under the marginal $p(t, \tau) = p(\tau|t)\pi(t)$ relieving the users of choosing σ manually. Details are derived in App. A.

3.2 Maximal Mode

In this mode, RoCUS finds trajectories that lead to maximal behavior values: $b^* \to \pm\infty$. It can also be used for finding minimal behavior values by negating the behavior. The posterior formulation is:

$$b_0 = \frac{b - \mathbb{E}[b]}{\sqrt{\mathbb{V}[b]}}, \quad \beta = \frac{1}{1 + e^{-b_0}}, \quad \widehat{\beta} \sim \mathcal{N}\left(\beta, \sigma^2\right), \quad t, \tau \sim p(t, \tau | \widehat{\beta} = 1), \tag{4}$$

where $\mathbb{E}[b]$ and $\mathbb{V}[b]$ are the mean and variance of the marginal $p(b)$. σ is chosen such that

$$\int_{1-\sqrt{3}\sigma}^{1} p(\beta)\,\mathrm{d}\beta = \alpha, \tag{5}$$

where $p(\beta)$ is the marginal distribution similar to Eq. 3. If $p(b)$ is normal, $p(\beta)$ is logit-normal. This formulation is again scale-invariant and has the same "volume" interpretation for α (App. A).

3.3 Posterior Sampling

Posterior sampling is different depending on the stochasticity of the controller and the dynamics.

Deterministic Controller and Dynamics: When both the controller and the dynamics are deterministic, so is $\tau|t$, denoted as $\tau(t)$. Eq. 2 reduces to $t \sim p(t|\widehat{b} = b^*) \propto p(\widehat{b} = b^*|t, \tau(t))\pi(t)$, and similarly for Eq. 4.

Using MH for sampling is now viable, as presented in Alg. 1. First, σ is computed from α (Line 2). Then we start with an initial task t (Line 3). For each of the N iterations, we propose a new task t_{new} according to a transition kernel and compute the forward and reverse transition probabilities $p_{\text{for}}, p_{\text{rev}}$ (Line 5). We evaluate the posteriors under t and t_{new} (Line 6 and 7) and calculate the acceptance ratio using the MH detailed balance principle (Line 8). Finally, we accept or reject accordingly (Line 9 – 11). Note that if the proposal is rejected, the current t is left unchanged *and appended to the samples*. In the end, we can discard the first N_B samples as burn-in, and/or thin the samples by only keeping every N_T samples to reduce auto-correlation.

Algorithm 1: MH Sampling Procedure

Input: "Posterior volume" α, number of samples N, optional burn-in N_B and thinning period N_T.

1 samples $\leftarrow$ [ ];
2 Get σ from α by Eq. 3 (matching) or 5 (maximal);
3 Randomly initialize t;
4 **for** $i = 1, ..., N$ **do**
5 $\quad t_{\text{new}}, p_{\text{for}}, p_{\text{rev}} = \text{propose}(t)$
6 $\quad$ Get p from t by Eq. 2 (match) or Eq. 4 (max)
7 $\quad$ Get p_{new} from t_{new} by Eq. 2 or Eq. 4;
8 $\quad a \leftarrow (p_{\text{new}} \cdot p_{\text{rev}})/(p \cdot p_{\text{for}})$;
9 $\quad$ Sample $u \sim \text{Unif}[0, 1]$;
10 $\quad$ **if** $u < a$ **then**
11 $\quad\quad t \leftarrow t_{\text{new}}$;
12 $\quad$ Append t to samples;
13 Optionally, discard the first N_B burn-in samples and thin the samples by only keeping every N_T samples;
14 **return** samples

Stochastic Controller: When the controller and $p(\tau|t)$ are stochastic, we need to sample in the combined (t, τ)-space. Typically the controller can be implemented by sampling a random variable u, and then producing the action based on the realization of u. For instance, for a Normal stochastic policy represented as action $a \sim \mathcal{N}(\mu(s), \sigma(s)^2)$ at state s, we can sample $u \sim \mathcal{N}(0, 1)$ and calculate $a = \mu(s) + u \cdot \sigma(s)$. Fig. 3 shows the graphical model diagram that explicitly accounts for the controller randomness with random variable u.

Using this transformation, Eq. 2 can be developed as

$$p(t, \tau|\widehat{b} = b^*) \propto p(\widehat{b} = b^*|t, \tau(e, u))p(u|t)\pi(t), \tag{6}$$

where we overload $\tau(t, u)$ to refer to the *deterministic* trajectory given the task t and controller randomness u. It is crucial that we can evaluate $p(u|t)$ for any u. Typically, we have independence $u \perp t$, so $p(u|t) = p(u)$. Concretely, building upon Alg. 1, u_{new} is proposed alongside with t_{new} (Line 5), the detailed balancing principle (Line 8) is multiplied by $p_{u,\text{rev}}/p_{u,\text{for}}$, and $t_{\text{new}}, u_{\text{new}}$ are accepted or rejected together (Line 10 – 12).

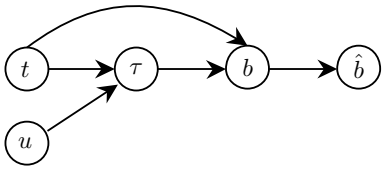

Figure 3: The same graphical model as in Fig. 2, but with the addition of stochasticity u in the controller such that $\tau|t, u$ is now deterministic.

Furthermore, this framework could also account for stochasticity in dynamics, but since it is not present in experiment, we discuss the necessary adaptations to RoCUS in App. B.

4 Behavior Taxonomy

Robot behaviors broadly belong to one of two classes: intentional and emergent. Behaviors that are specified as objectives for a controller are *intentional*. For example, in a reaching problem a controller likely optimizes to move the end-effector to the target. This optimization occurs by setting the target as an attractor in a DS, setting the objective configuration so the target is reached in RRT,

or rewarding proximity in RL. Final distance between the end-effector and the target is an intentional behavior for all three controllers. In comparison, *emergent* behaviors are not explicitly specified in the objective. For this same reaching problem, an RL policy with reward based solely on distance may exhibit smooth trajectories for some target locations and jerky, abrupt trajectories for others. Such unspecified behaviors may emerge due to robot kinematic structure, training stochasticity, or model inductive bias.

While many behaviors are specific to robots or tasks, some are widely applicable to different robots (e.g., arm, humanoid, or soft robots). Roboticists applying these metrics must choose an appropriate subset and specific instantiations for their robot morphology and task. Given a trajectory τ, most behavior metrics $b(\tau, t)$ can be expressed as a line integral of a scalar field $V(\mathbf{x})$ along τ or its length-normalized integral,

$$b = \int_{\tau} V(\mathbf{x})\, \mathrm{d}s \quad \text{or} \quad b = \frac{1}{||\tau||} \int_{\tau} V(\mathbf{x})\, \mathrm{d}s, \tag{7}$$

where $\mathrm{d}s$ is the infinitesimal segment on τ at $\mathbf{x}$ and $||\tau||$ is the trajectory length. $\mathbf{x}$ and τ can be in either joint space or task space. We numerically compute it via Riemann integration.

We introduce five classes of behaviors: *length*, *time derivatives* (i.e. velocity, acceleration and jerk), *straight-line deviation*, *obstacle clearance*, and *motion legibility*, as detailed in App. C. Users also can add alternative behaviors for use with RoCUS.

5 RoCUS Use Case Demos

In this section, we concretely demonstrate how RoCUS can be used to find "hidden" properties of various controllers for two common tasks, navigation and reaching. Furthermore, during the analyses we uncovered a suboptimal controller design due to bad hyper-parameter choices, and findings by RoCUS point out a remedy option, which results in a much better controller.

5.1 Controller Algorithms

In total, we consider four classes of robot controllers. The **imitation learning** (IL) controller uses expert demonstrations to learn a neural network policy which maps observations to deterministic actions. The **reinforcement learning** (RL) controller implements proximal policy gradient (PPO) [27]. While a mean and a variance is used to parameterize a PPO policy, at execution time, the policy deterministically outputs the mean action. The **dynamical system** (DS) controller modulates the linear controller $\mathbf{u}(\mathbf{x}) = \mathbf{x}^* - \mathbf{x}$, for the task-space target $\mathbf{x}^*$, using the modulation matrix M proposed by [7]: $\mathbf{u}_M(\mathbf{x}) = M \cdot \mathbf{u}(\mathbf{x})$. M is calculated from a list of monotonically increasing Γ-functions, one per obstacle. We give a self-contained review of this approach in App. D.

The **rapidly-exploring random tree** (RRT) controller finds a configuration-space trajectory via RRT and then controls the robot through descretized segments. Unlike the previous two, RRT is a stochastic controller, and we discuss the use of controller stochasticity u (c.f. Fig. 3) in App. E.

The MCMC sampling uses a Gaussian drift kernel, and additional details are in App. F.

5.2 2D Navigation Task Experiments

Setup In a rectangular arena with irregularly shaped obstacles, a point mass robot needs to move from the lower left to the upper right corner (Fig. 1 left). App. G details the obstacle generation and robot simulation procedures and contains more environment visualizations.

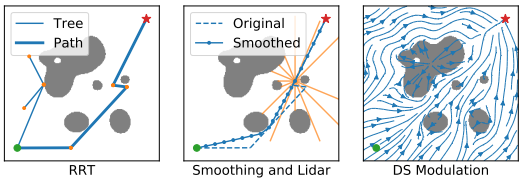

Figure 4: RRT, IL and DS controllers on 2D navigation domain. Left: the RRT controller tree. Middle: smoothed RRT trajectory and lidar sensor (orange lines) for IL controller training. Right: the modulation by the DS controller.

We consider three controllers for this environment: an RRT planner, a deep learning IL policy, and a DS (Fig. 4). The RRT planner implements Alg. 2 and discretizes the path to small segments as control signals at each time step. The IL controller uses smoothed RRT trajectories as expert demonstrations, and learns to predict heading angle from its current position and lidar readings. The DS controller finds an interior reference point for each obstacle, and converts each obstacle in the environment to be star-shaped. Γ-functions are then defined for these obstacles and used to compute the modulation matrix M. App. H contains additional implementation details.

Straight-Line Deviation In most cases, the robot cannot navigate directly to the target in a straight line. Thus, the collision-avoidance behavior is a crucial aspect for for navigation robots. To under-

stand it, we sample obstacles that lead to trajectories minimally deviating from the straight line path. Since the deviation is always non-negative, we use the match-type posterior in Eq. 2 with target $b^* = 0$. In Fig. 5, the top row plots posterior trajectories in orange, with prior trajectories in blue, for three controllers. The bottom row plots the obstacle distributions compared to the prior, with red regions being more likely to be occupied by obstacles and blue ones less likely to be obstructed.

The visualization shows that for DS and RRT, the posterior trajectories and obstacle configurations are mostly symmetric with respect to the straight-line connection, as expected since both methods are formulated symmetrically with respect to the x- and y-coordinates.

For the obstacle distribution, the RRT result is expected. Since RRT seeks straight-line connections whenever possible, the posterior largely favor a "diagonal corridor" with obstacles on either side. For the DS, obstacles are slightly *more* likely to exist at the two ends of the above-mentioned corridor. This behavior is an artifact of the DS *tail effect*, which drags the robot around the obstacle (details in App. D). By taking advantage of anchor-like obstacles at the ends of the corridor, the modulation can reliably minimize the straight-line deviation.

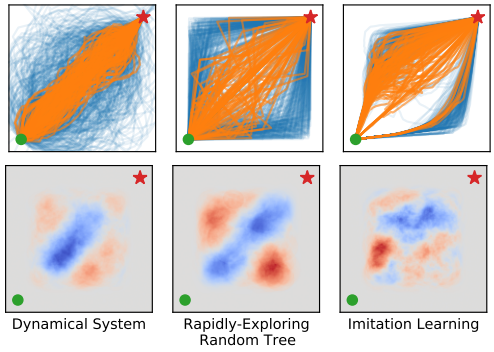

Figure 5: Top: Posterior trajectories in orange vs. prior in blue for minimal straight-line deviation behavior. Bottom: Posterior obstacle distribution, with red for higher obstacle density and blue for lower. IL is notably asymmetric, and DS notably has obstacles obscuring the direct path.

The IL controller saliently exhibits trajectory asymmetry: it mostly takes paths on the left of the robot. It is possible that the asymmetry is due to "unlucky" samples by the MH sampler, but many independent restarts all confirm the asymmetry, indicating that the asymmetry is inherent in the learned model. Since the neural network architecture is symmetric, we conclude that the stochasticity in the dataset generation and training procedure (e.g. initialization) leads to such imbalanced behaviors. Furthermore, the obstacle map suggests that obstacles are distributed very close to the robot path. Why does the robot seem to drive into obstacles? The answer lies in dataset generation process: the smoothing procedure (Fig. 4 middle) results in most demonstrated paths navigating tightly around obstacles, and it is therefore natural to expect the learned IL controller to display the same behavior.

Takeaways RoCUS reveals two unexpected phenomena. First, IL trajectories are highly asymmetric toward the left of the obstacle due to dataset and/or training imbalance. Second, both DS and IL models exhibit certain "obstacle-seeking" behaviors, the former due to the "tail-effect" and the latter due the dataset generation process. In both cases, such behavior may be undesirable in deployment due to possibly imprecise actuation, and the controller design may need to be modified. We also studied the legibility and obstacle clearance of these controllers, with results in App. I.

5.3 7DoF Arm Reaching Task Experiments

Setup A 7DoF Franka Panda arm is mounted on the side of a table with a T-shaped divider (Fig. 1 right). Starting from the same initial configuration on top of the table, it needs to reach a random location on either side under the divider. We simulate this task in PyBullet [28]. We again consider three controllers: an RRT planner, a deep RL PPO agent, and a DS formulation.

RRT again implements Algorithm 2, but uses inverse kinematics (IK) to first find the joint configuration corresponding to the target location. The RL controller is a multi-layer perceptron (MLP) network trained using the PPO algorithm. The DS model outputs the end-effector trajectory in the task space, which is converted to joint space via IK, with SVM-learned obstacle definitions. Overall, RRT and RL are quite successful in reaching the target while the DS is not due to the bulky robot structure, close proximity to the divider, and the task-space only modulation. App. J contains additional implementation details for each method.

End-Effector Movement We find configurations that minimize the total travel distance of the end-effector for RRT and RL (DS omitted due to high failure rate). Fig. 6 (left two) shows the posterior target locations and trajectories. Notably, unlike RL, RRT trajectories are highly asymmetric, since there are straight-line connections in the configuration space from the initial pose to some target regions on the left, while every right-side goal requires at least an intermediate node.

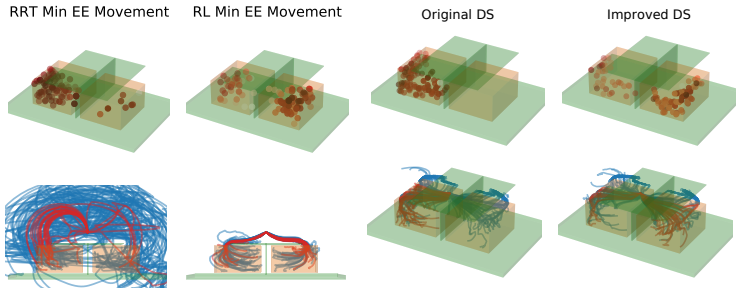

Figure 6: Left: Minimal end-effector movement samples for RRT and RL. Right: Posterior samples for minimal distance from end-effector to target for the original and improved DS controllers. Top: posterior targets locations, with tabletop + divider in green and target region in orange. Bottom: posterior trajectories in red, prior trajectories in blue. Robot is mounted on the near long edge.

DS Improvement with RoCUS Our initial DS implementation frequently failed to reach the target. This is understandable, as the DS convergence guarantee [7] is only valid in task space, in which we defined the modulation. When the full-arm motion is solved via IK, it is possible that some body parts may collide and get stuck because of the table divider. To understand the DS behaviors, we used RoCUS to sample target locations that result in minimal final distance from the end-effector to the target (i.e., most successful executions, Fig. 6 center-right). Similar to the RRT case, the samples show strong lateral asymmetry, with all posterior target locations on the left, due to the same cause of asymmetric kinematic structure. The result points to a clear path to improve the DS controller such that it can succeed with right-side targets: increase the collision clearance of the divider so that the end-effector navigates farther away from the divider, thus also bringing the whole arm to be farther away. As detailed in App. K.1, this modification greatly improves the controller performance as confirmed by the new symmetry in Fig. 6 (rightmost). In addition, since the issue with DS controller mainly lies in obstacle avoidance in joint-space or on the body of the robot, additional techniques [29, 30, 31] could be used and we leave them to future directions.

Takeaway The set of studies reveal an important implication of the robot's kinematic structure: the left side is much less "congested" with obstacles than the right side in the configuration space. While the RL method is able to learn efficient policies for both sides, the design of certain controllers may need to explicitly consider such factors. An additional study on legibility is in App. K.

5.4 Quantitative Summary

We performed additional studies of other behaviors on both tasks, and present quantitative summary of the prior vs. posterior mean behavior values in Table. 1. We can see that RoCUS can consistently find samples that are salient with respect to the target behavior.

Domain	Behavior	Target	Prior (DS)	Post. (DS)	Prior (IL/RL)	Post. (IL/RL)	Prior (RRT)	Post. (RRT)
2D Nav	Avg. Jerk	0	1.84e-3	1.46e-3	6.95e-4	3.19e-4	4.24e-4	2.79e-4
	Straight	0	0.256	0.084	0.378	0.301	0.470	0.162
	Legibility	min	0.819	0.650	0.877	0.784	0.798	0.669
	Obstacle	0	0.309	0.229	0.262	0.218	0.312	0.241
	Obstacle	max	0.309	0.611	0.262	0.387	0.312	0.442
Arm	Straight	0	0.980	0.913	0.858	0.762	1.223	0.897
	EE Dist	0	0.934	0.623	0.958	0.691	3.741	1.192

Table 1: Quantitative results on additional tasks for two domains.

6 A Deeper Look at the Sampling Procedure (New Section)

After confirming that RoCUS can indeed uncover significant and actionable insights into various controllers, we study the sampling procedure itself, using various tasks described above as examples.

Mixing Property A potential downside of MCMC sampler is the slow mixing time, causing the chain to take a long time to converge from initialization and causing consecutive samples to be highly correlated with each other. Does this phenomenon happen for our sampling? Fig. 7 plots the behavior value for the MCMC chains of a sampling task, showing that the chain mixes well quite fast (additional ones in Fig. 10 of App. F). Thus, we could reassure ourselves that a modest amount of samples, such as several thousand, is typically sufficient to model the target posterior distribution well.

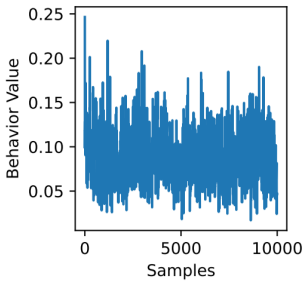

Figure 7: MCMC chain.

Baseline: Top-k Selection To the best of our knowledge, RoCUS is the first work that applies the principle of transparency-by-

example [12] to robotic tasks, and we are not aware of existing methods for the same purpose. Notably, adversarial perturbation algorithms [5] are *not* feasible, since the simulator (or the real world) stepping are not typically differentiable.

As discussed in Sec. 1, a straightforward alternative to RoCUS is to simply run the controller on N different scenarios and pick the top-k with respect to the target behavior. We demonstrate the shortcomings of this approach on the minimal straight-line deviation behavior for the 2D navigation DS controller (RoCUS samples shown in Fig. 5 left). Fig. 8 (left) shows the trajectories of different values of k for the same fixed N, and vice versa.

While a bigger N/k ratio leads to more salient behaviors in the top-k samples, these examples become more concentrated around the global maximum and less diverse, making this approach especially myopic. Further, it is not easy to find the optimal N to trade off between diversity and saliency of the top-k samples. By contrast, RoCUS offers the intuitive α hyper-parameter. The two middle panels shows that a smaller N fails to highlight the "corridor" pattern while a larger N makes it completely open and fails to show the "tail-effect anchors" at the two ends.

In addition, the hard cut-off at the k-th salient behavior threshold has two undesirable implications: first, every trajectory more salient than the threshold is kept but is given equal importance; second, a trajectory even slightly under the threshold is strictly discarded. By comparison, RoCUS gives more importance to more salient samples in a progressive manner, as shown in Fig. 8 right.

Finally, top-k selection is very computationally inefficient. It requires all the unselected $N - k$ samples to be discarded, while RoCUS is much more efficient in that all samples after the burn-in can be kept (with no thinning) since the posterior concentrated on the salient behavior is directly sampled.

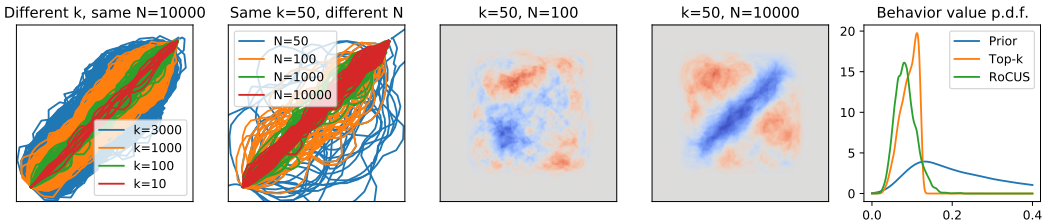

Figure 8: Top-k selection baseline. Left two: trajectory distribution; middle two: obstacle distribution; right one: probability density function of behavior values.

7 Discussion and Future Work

RoCUS enables humans to build better mental models of robot controllers. It complements existing task-completion-centric evaluations on hand-designed scenarios by generating scenarios and trajectories for specified behaviors in a principled way. We demonstrated how it can uncover non-obvious insights in two domains, and showed how it can help with debugging and improving a controller.

While RoCUS is mainly a tool to analyze robot controllers in simulation as part of comprehensive testing before deployment, it can help understanding of real world behaviors. When an anomaly is observed during real world deployment, the developers can use RoCUS to sample the anomalous aspect of the behavior, and see if the robot behaves similarly in other cases, which might indicate a pattern of systematic failures. Furthermore, it is not inherently limited to simulation: the only requirement is to be able to reset the environment to a specific task. For the arm reaching task, this is trivial, as there are no movable objects or obstacles. For autonomous driving, "recreating" a traffic condition that involves other vehicles may be hard. However, a key advantage of RoCUS is the decoupling of the task description and the controller algorithm, which allows testing on simpler task variants (e.g. with props instead of real cars) in complicated scenarios.

There are multiple directions for future work, including evaluation of *model updates* [32] by defining behavior functions on two controllers, better understanding the samples with explainable artificial intelligence (XAI), and an appropriate interface to facilitate the two-way communication between RoCUS and end-users, as discussed in detail in App. L.

Overall, RoCUS is a framework for systematic discovery and inspection of representative robotic controller scenarios. In general, we hope the demonstrated usefulness of tools like RoCUS sparks further efforts towards developing support for more holistic understanding of robot controllers.

References

[1] J. Bort. Inside Uber before its self-driving car killed a pedestrian: Sources describe infighting, 'perverse' incentives, and questionable decisions, 2018. URL https://www.businessinsider.com/sources-describe-questionable-decisions-and-dysfunction-inside-ubers-self-driving-unit-before-one-of-its-cars-killed-a-pedestrian-2018-10.

[2] A. Rasouli and J. K. Tsotsos. Autonomous vehicles that interact with pedestrians: A survey of theory and practice. *IEEE transactions on intelligent transportation systems*, 21(3), 2019.

[3] A. Bobu, A. Bajcsy, J. F. Fisac, S. Deglurkar, and A. D. Dragan. Quantifying hypothesis space misspecification in learning from human–robot demonstrations and physical corrections. *IEEE Transactions on Robotics*, 36(3):835–854, 2020.

[4] G. J. Myers, T. Badgett, T. M. Thomas, and C. Sandler. *The art of software testing*, volume 2. Wiley Online Library, 2004.

[5] I. J. Goodfellow, J. Shlens, and C. Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014.

[6] B. D. Argall, S. Chernova, M. Veloso, and B. Browning. A survey of robot learning from demonstration. *RAS*, 57(5):469–483, 2009.

[7] L. Huber, A. Billard, and J.-J. Slotine. Avoidance of convex and concave obstacles with convergence ensured through contraction. *IEEE Robotics and Automation Letters*, 2019.

[8] S. M. LaValle. *Planning algorithms*. Cambridge university press, 2006.

[9] R. S. Sutton, A. G. Barto, et al. *Introduction to reinforcement learning*, volume 135. MIT press Cambridge, 1998.

[10] M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I. D. Raji, and T. Gebru. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 220–229, 2019.

[11] M. T. Ribeiro, T. Wu, C. Guestrin, and S. Singh. Beyond accuracy: Behavioral testing of nlp models with checklist. *arXiv preprint arXiv:2005.04118*, 2020.

[12] S. Booth, Y. Zhou, A. Shah, and J. Shah. Bayes-TrEx: a bayesian sampling approach to model transparency by example. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021.

[13] C. Fan, X. Qin, and J. Deshmukh. Parameter searching and partition with probabilistic coverage guarantees. *arXiv preprint arXiv:2004.00279*, 2020.

[14] B. Hayes and J. A. Shah. Improving robot controller transparency through autonomous policy explanation. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, pages 303–312. IEEE, 2017.

[15] I. Lage, D. Lifschitz, F. Doshi-Velez, and O. Amir. Exploring computational user models for agent policy summarization. *arXiv preprint arXiv:1905.13271*, 2019.

[16] I. Rahwan, M. Cebrian, N. Obradovich, J. Bongard, J.-F. Bonnefon, C. Breazeal, J. W. Crandall, N. A. Christakis, I. D. Couzin, M. O. Jackson, et al. Machine behaviour. *Nature*, 2019.

[17] J. Mahler, R. Platt, A. Rodriguez, M. Ciocarlie, A. Dollar, R. Detry, M. A. Roa, H. Yanco, A. Norton, J. Falco, et al. Guest editorial open discussion of robot grasping benchmarks, protocols, and metrics. *TASE*, 15(4):1440–1442, 2018.

[18] A. Murali, T. Chen, K. V. Alwala, D. Gandhi, L. Pinto, S. Gupta, and A. Gupta. Pyrobot: An open-source robotics framework for research and benchmarking. 2019.

[19] S. James, Z. Ma, D. R. Arrojo, and A. J. Davison. RLBench: The robot learning benchmark & learning environment. *IEEE Robotics and Automation Letters*, 5(2):3019–3026, 2020.

[20] P. Anderson, A. Chang, D. S. Chaplot, A. Dosovitskiy, S. Gupta, V. Koltun, J. Kosecka, J. Malik, R. Mottaghi, M. Savva, et al. On evaluation of embodied navigation agents. 2018.

[21] B. Cohen, I. A. Şucan, and S. Chitta. A generic infrastructure for benchmarking motion planners. In *IROS*, pages 589–595. IEEE, 2012.

[22] M. Moll, I. A. Sucan, and L. E. Kavraki. Benchmarking motion planning algorithms: An extensible infrastructure for analysis and visualization. *IEEE RAM*, 22(3):96–102, 2015.

[23] F. Lagriffoul, N. T. Dantam, C. Garrett, A. Akbari, S. Srivastava, and L. E. Kavraki. Platform-independent benchmarks for task and motion planning. *IEEE Robotics and Automation Letters*, 3(4):3765–3772, 2018.

[24] A. Lemme, Y. Meirovitch, M. Khansari-Zadeh, T. Flash, A. Billard, and J. J. Steil. Open-source benchmarking for learned reaching motion generation in robotics. *Paladyn, Journal of Behavioral Robotics*, 1(open-issue), 2015.

[25] A. D. Dragan, K. C. Lee, and S. S. Srinivasa. Legibility and predictability of robot motion. In *HRI*. IEEE, 2013.

[26] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. 1970.

[27] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[28] E. Coumans and Y. Bai. Pybullet, a python module for physics simulation for games, robotics and machine learning. 2016.

[29] O. Khatib. Real-time obstacle avoidance for manipulators and mobile robots. In *Autonomous robot vehicles*, pages 396–404. Springer, 1986.

[30] N. D. Ratliff, J. Issac, D. Kappler, S. Birchfield, and D. Fox. Riemannian motion policies. *arXiv preprint arXiv:1801.02854*, 2018.

[31] J. Urain, A. Li, P. Liu, C. D'Eramo, and J. Peters. Composable energy policies for reactive motion generation and reinforcement learning. *arXiv preprint arXiv:2105.04962*, 2021.

[32] G. Bansal, B. Nushi, E. Kamar, D. S. Weld, W. S. Lasecki, and E. Horvitz. Updates in human-ai teams: Understanding and addressing the performance/compatibility tradeoff. In *AAAI*, 2019.

[33] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng. *Handbook of markov chain monte carlo*. CRC press, 2011.

[34] S. M. Khansari-Zadeh and A. Billard. A dynamical system approach to realtime obstacle avoidance. *Autonomous Robots*, 32(4):433–454, 2012.

[35] Y. Zhou and K. Hauser. 6DOF grasp planning by optimizing a deep learning scoring function. In *RSS Workshop on Revisiting Contact – Turning a Problem into a Solution*, 2017.

[36] H. Xu, Y. Gao, F. Yu, and T. Darrell. End-to-end learning of driving models from large-scale video datasets. In *CVPR*, pages 2174–2182, 2017.

[37] K. Hauser. Robust contact generation for robot simulation with unstructured meshes. In *IJRR*. Springer, 2016.

[38] S. S. Mirrazavi Salehian, N. Figueroa, and A. Billard. A unified framework for coordinated multi-arm motion planning. *The International Journal of Robotics Research*, 37(10):1205–1232, 2018.

[39] S. Karaman and E. Frazzoli. Sampling-based algorithms for optimal motion planning. *IJRR*, 30(7):846–894, 2011.

[40] K. Hauser and Y. Zhou. Asymptotically optimal planning by feasible kinodynamic planning in a state–cost space. *TRO*, 32(6):1431–1443, 2016.

[41] S. Greydanus, A. Koul, J. Dodge, and A. Fern. Visualizing and understanding atari agents. In *ICML*, pages 1792–1801. PMLR, 2018.

[42] T. Zahavy, N. Ben-Zrihem, and S. Mannor. Graying the black box: Understanding dqns. In *ICML*, pages 1899–1908. PMLR, 2016.

[43] A. Atrey, K. Clary, and D. Jensen. Exploratory not explanatory: Counterfactual analysis of saliency maps for deep reinforcement learning. *arXiv preprint arXiv:1912.05743*, 2019.