Infrastructure to support robots: a practical, scalable model for comparative evaluation of design choices

Grace McFassel Dylan A. Shell

Abstract— **It is easier to program effective robots when they inhabit highly structured environments. The growing literature on methods to aid robot design has given comparatively little consideration to elements external to the robot itself, yet such elements can encode or enhance information (to improve perception), can alter the effects or costs of actions (to help control), and can provide regularity by imposing constraints. External elements have the potential to be shared, to scale elastically, and to spread both benefits and installation/operating costs. These are traits of *infrastructure in support of robots*. We introduce a basic but flexible mathematical model—via the MDP framework—for rational evaluation of proposed additions and changes to environments, including where infrastructure may improve precision or performance of either perception or actuation. Through it, one can assess the numbers of agents needed for infrastructure investment to be economical, determine when installation costs would be recouped, and evaluate the effect of behavior changes as responses to environmental modifications. To demonstrate how the model can be instantiated, four simple but practical case studies are presented.**

I. INTRODUCTION

A growing body of work proposes computational approaches to robot design, including research on the selection/optimization of actuators and sensors that meet some desired level of performance while balancing cost and efficiency [1]–[4]. Such work considers elements that are part of, and physically internal to, the robot itself. While it is generally understood that structured environments ease many of the challenges involved in developing and deploying useful robots, this paper approaches the idea with a fresh twist by treating this within a design problem.

Consider tug vehicles deployed in a specially instrumented fulfillment center. It isn't entirely obvious for such a system where the boundary of the "robot" ought to be:— aside from the mobile tug shunting packages around, the instruments and facility itself are key to the vehicle's efficacy. This paper introduces a method for making informed design choices involving infrastructure *for* robots; the focus is on environmental elements possessing infrastructure-like attributes, which is not solely a question of physical (external vs internal) placement. First, the nebulous concept of infrastructure will be clarified, including the distillation of common traits that lead to important design questions and considerations.

A. Informal definitions and paper contributions

Infrastructure is commonly used to describe a variety of services and projects that are made available to large numbers of users. Here, we aim to informally identify the traits that are often common across different kinds of infrastructure, which will provide this paper's working definition. Infrastructure that we examine has these six features:

- ▸ **Group Utilization** Multiple agents are able to access the infrastructure to benefit from it; e.g. roadways, satellites.
- ▸ **Elastic Scaling** It should be capable of supporting the intended number of users and of being extended in the future.
- ▸ **Reusable** Infrastructure should endure multiple uses before being consumed, repaired, etc. and not be perishable.
- ▸ **Cost Distribution** Recuperation of the upfront construction and maintenance costs are distributed over the users in some way; e.g. taxes, tolls, monthly bills.
- ▸ **Fairness** It should not harm any one group unduly.
- ▸ **Impacts Agent Behavior** Finally, we expect that the infrastructure should alter the operation of agents in the environment, having some measurable impact.

Under this definition, we aim to answer the following: How can different proposed infrastructure be compared? Can we examine how infrastructure will impact large populations of agents without exhaustive simulation? Assessment should include some subset of robots altering their behavior to suit new infrastructure. What is the impact on agents with differing abilities and goals, and how can that impact be translated into a concept of "fairness"? How can the cost of infrastructure be compared to benefits in performance?

To formalize the above description, we propose a model that treats robots via Markov Decision Processes (MDPs); the approach prioritizes practicality by being simple and flexible. The infrastructure perspective offers a subtly different approach to improving robot performance from standard methods. The paper's later sections discuss specific instances of infrastructure and demonstrate its promise. Note, finally, that a fuller treatment does appear in [5, Ch. 4].

II. RELATED WORK

Most work on robot design does not consider the robot as a part embedded in a much larger enveloping system, itself amenable to design; a recent and notable exception is [6]. A long line of robotics research explored *stigmergic* multi-robot teams, wherein coordination is mediated through modifications of the shared environment [7], often in task-directed ways [8], [9]. This perspective, emphasizing the environment as something active or able to be exploited structurally (hosting common markers, or persistent shared computation) rather than merely being passive circumambient space, leads to new ways to coordinate robots [10]–[13].

Setting up systems that shape (or constrain) the behavior of agents who make decisions autonomously falls within the purview of economics. The specific economic theory of

Both authors are affiliated with Dept. of Computer Science & Engineering, Texas A&M University, College Station, TX, USA. {gracem | dshell}@tamu.edu. This work was supported in part by the National Science Foundation under Grant No. [IIS-2034097] and a graduate fellowship provided by the 3M Company and 3M Gives.

clubs, in which members all derive a benefit from goods while dividing costs [14], fits infrastructure well; it can describe how the transition of a good owned by a single agent to a shared one may change its utility [15]. Such work is connected with equilibria and game theory [16], the latter having been given serious consideration as a potential unifying formal model for robotics [17]. Algorithmic mechanism design [18] drawing on economic ideas, was spurred on initially by routing games as models of behavior on networks. Traffic networks are a very visible example of infrastructure, with models going back to the 1950s [19] and increasing use of agent-based techniques [20]. The six infrastructure features/traits listed on the previous page are conceptually broader than is typical in robotic design; the importance of such broadening is championed by [21] in the context of fleets of autonomous vehicles.

Reinforcement learning has been applied to traffic networks that modify infrastructure to influence the behavior of other agents. Demonstrations change the pathing and behavior of individual agents [22] as well as analyzing and responding to general trends such as congestion [23]. Fundamentally, reinforcement learning examines how agents can operate (and adapt to changes in) an unknown environment; the standard setting uses MDPs with a policy evolving over time [24]. Both the MDP model and changes to policies appear in our work, but we consider the impact of environment change after policies have converged, not the transient phase progressing toward convergence.

III. PROBLEM FORMALIZATION

Our treatment is intended to apply broadly and this generality demands that the presentation be quite abstract.

A. Preliminaries

We model each agent's interactions with its environment, indicated as M, as a Goal MDP, which we recall next:

Definition 1 (Goal MDP [24])**.** A Goal Markov Decision Process, or MDP, consists of:
- a finite set of states S, with an initial state[1] $s_0 \in S$;
- a finite set of actions A;
- a transition model $T : S \times S \times A \to [0, 1]$ such that $T(s', s, a) = \Pr(s'|s, a)$ denotes the probability of arriving in state s' having issued action a from state s;
- a function $C : S \times A \times S \to \mathbb{R} \cup \{+\infty\}$ where $C(s, a, s')$ is the expected cost expended for the agent occupying state s taking action a and arriving at state s';
- a nonempty selection of goal states $G \subseteq S$.

Agents' behavior will be described via a policy:

Definition 2 (Policy)**.** A policy $\pi : S \to A$ assigns, for each state $s \in S$, an action $\pi(s) \in A$ for an agent to take.

While we will discuss optimal policies in Section IV for comparative purposes, the ensuing definitions apply to all potential policies for an MDP.

[1]While we will assume s_0 is a single state, generalization to a distribution of starting states, as well as instances where the starting configuration is itself part of the optimization problem, is straightforward.

Definition 3 (Expected Cumulative Cost)**.** For $M = (S, s_0, A, T, C, G)$, the *expected cumulative cost* of policy π is

$$E[\pi|M] = \mathop{\mathbb{E}}_{s_0, s_1, \ldots, s_m} \left(\sum_{i=0}^{m} C(s_i) \right),$$

where the expectation is over finite sequences $s_0, s_1, \ldots, s_m$ arriving at some goal state $s_m \in G$, with probabilities $T(s_i, s_{i-1}, \pi(s_{i-1}))$, for all i. If there are no such sequences with non-zero probability, then we declare $E[\pi|W] = +\infty$.

B. Defining and Applying Infrastructure

We distinguish $K \in \mathbb{N}$ different *classes* (or types) of agent.

Definition 4 (Environment)**.** For K classes of agent, an *environment* $\mathcal{E}$ is a collection of MDPs, one for each class:

$$\mathcal{E} := \{M_1, M_2, M_3, \ldots, M_K\}.$$

To evaluate some proposed piece of infrastructure, each MDP must be modified to "apply" its effects. This is formalized as a transformation on environments:

Definition 5 (Infrastructure)**.** Collection $\mathbb{I}$ of K triples is *infrastructure* if we have $\mathbb{I} := \{(h_1^{\mathrm{T}}, h_1^{\mathrm{C}}, c_1), \ldots, (h_K^{\mathrm{T}}, h_K^{\mathrm{C}}, c_K)\}$, each triple $(h_i^{\mathrm{T}}, h_i^{\mathrm{C}}, c_i)$ comprising two maps and a scalar:
- function h_i^{T} mapping from the original transition function of MDP M_i to a new transition function T':
$$T_{M_i}(\cdot,\cdot,\cdot) \xmapsto{h_i^{\mathrm{T}}} T'_{M_i'}(\cdot,\cdot,\cdot);$$
- function h_i^{C} mapping from the original cost function C of MDP M_i to a new cost function C':
$$C_{M_i}(\cdot,\cdot,\cdot) \xmapsto{h_i^{\mathrm{C}}} C'_{M_i'}(\cdot,\cdot,\cdot); \text{ and}$$
- $c_i \in \mathbb{R}^{\geq 0}$, an associated construction cost.

For each i, the pair h_i^{T} and h_i^{C} modify MDP M_i, altering its transition and cost functions, to give the new MDP M_i'. Then, infrastructure $\mathbb{I}$ is an operator that modifies some environment $\mathcal{E} \xmapsto{\mathbb{I}} \mathbb{I}(\mathcal{E}) = \{M_1', M_2', M_3', \ldots, M_K'\}$.

In the above definition, each triple $(h_i^{\mathrm{T}}, h_i^{\mathrm{C}}, c_i)$ contains a construction cost c_i, which represents the one-time cost of constructing the infrastructure, not to be confused with the cost functions of the MDPs within $\mathcal{E}$.[2] While here we consider only the initial price of construction, much infrastructure requires continual upkeep and consideration of maintenance costs incurred over time is the topic of Section V. The benefit provided by this infrastructure's construction relies on several factors: the number and types of agents within $\mathcal{E}$, if (and how) agents change their behavior in the presence of infrastructure, and how this affects expected costs. To formalize these aspects, we start with the fact that environments are inhabited by (typically multiple) agents:

Definition 6 (Agent population)**.** An *agent population* $\mathbf{P}$ for environment $\mathcal{E} = \{M_1, M_2, \ldots, M_K\}$ is a collection of sub-populations, each representing a collection of a (fixed)

[2]We require that each MDP's costs and infrastructure cost c_i be in the same units — in our examples, the MDP's cost is converted to dollars so we may look at recouped costs. Determining the equivalent "worth" of a cost function is not always straightforward, but picking other units may help.

number of agents of a particular class, along with a policy describing their behavior:

$$\mathbf{P} \coloneqq \{\mathscr{P}_1, \mathscr{P}_2, \ldots, \mathscr{P}_{|\mathbf{P}|}\},$$

where each sub-population $\mathscr{P}_i \coloneqq (n_i, c_i, \pi_i)$ has $n_i \in \mathbb{N}$ agents of class $c_i \in \{1, \ldots, K\}$, whose behavior is modeled as following policy $\pi_i : S(M_{c_i}) \to A(M_{c_i})$.

(In the preceding, we have wrtten $S(M)$ for the states of MDP M; also, analogously, $A(M)$ for actions.) The connection between environments —including those modified by forms of infrastructure— and populations is captured next:

Definition 7 (Infrastructure response)**.** For environment $\mathcal{E}$, an *infrastructure response* is some rule that takes $\mathbb{I}$ and a given agent population

$$\mathbf{P} = \big\{(n_1, c_1, \pi_1), (n_2, c_2, \pi_2), \ldots, (n_{|\mathbf{P}|}, c_{|\mathbf{P}|}, \pi_{|\mathbf{P}|})\big\}$$

and produces

$$\mathbf{P}' = \big\{(n'_1, c'_1, \pi'_1), (n'_2, c'_2, \pi'_2), \ldots, (n'_{|\mathbf{P}'|}, c'_{|\mathbf{P}'|}, \pi'_{|\mathbf{P}'|})\big\}$$

such that $\sum\limits_{(n_i, k, \pi_i) \in \mathbf{P}} n_i = \sum\limits_{(n'_i, k, \pi'_i) \in \mathbf{P}'} n'_i$, for each $k \in \{1, \ldots, K\}$.

The intuition is that an infrastructure response reflects a change in a population where there may be a different apportioning of sub-populations, but where the total agents of each class is preserved. Clearly, also, the total number of agents in the population is conserved.

Next, to make this more tangible, we provide some concrete examples of infrastructure responses:

- ▷ An *oblivious utilization* is the identity infrastructure response $\mathbf{P} \mapsto \mathbf{P}$ regardless of infrastructure $\mathbb{I}$.
- ▷ Given any operator $\mathbf{S}(\cdot)$ that produces a policy from an MDP,[3] an $\mathbf{S}$-*based fully adaptive utilization* is the infrastructure response $\mathbf{P} \mapsto \mathbf{P}'$ where each $(n_i, c_i, \pi_i) \in \mathbf{P}$ becomes $(n_i, c_i, \pi'_i) \in \mathbf{P}'$ where π'_i is a policy obtained via $\mathbf{S}(M'_i)$, assuming $\mathbb{I}(\mathcal{E}) = \{M'_1, M'_2, M'_3, \ldots, M'_K\}$.
- ▷ Again using $\mathbf{S}(\cdot)$, for adoption rate $\alpha \in [0, 1]$, the *adoption-based utilization* is the infrastructure response $\mathbf{P} \mapsto \mathbf{P}'$ with every $(n_i, c_i, \pi_i) \in \mathbf{P}$ contributing two elements to $\mathbf{P}'$:
  1) $(n_i - \lfloor \alpha \cdot n_i \rfloor, c_i, \pi_i) \in \mathbf{P}'$, and
  2) $(\lfloor \alpha \cdot n_i \rfloor, c_i, \pi'_i) \in \mathbf{P}'$, where π'_i is obtained via $\mathbf{S}(M'_i)$, again assuming $\mathbb{I}(\mathcal{E}) = \{M'_1, M'_2, \ldots, M'_K\}$.

For the final case, when the adoption rate $\alpha = 0$ or $\alpha = 1$, one recovers the two previous instances.[4]

As the term "infrastructure response" connotes, these are representations of how populations of agents react to changes made to the world. Agents may not be aware of a change in the environment, resulting in oblivious utilization. (We will see in Section IV-A that even oblivious agents may see benefits from infrastructure.) Conversely, $\mathbf{S}$-based fully adaptive utilization is where all agents within the population update their policies according to the infrastructure transformation.

An agent can operate obliviously on the transformed MDP M' owing to the way in which infrastructure is defined. The transformation function h^{T} cannot eliminate any actions available to the agent at that state, although it may change outcomes. (This follows naturally as infrastructure is only an operation on the world, not on the agent's capabilities.) Thus, while there is no guarantee that an oblivious agent will succeed, the original policy π can be used on M'.

The notion of an adoption rate α is used in our analysis of adoption-based utilization to indicate what proportion of a class of agents create an updated policy π'. It models situations in which either some agents remain unaware of the modified environment, or, if aware, choose not to alter their behavior. The realization splits a single sub-population into two groups, one with the identity infrastructure response while the other generates a new policy.

With these elements rigorously defined, we next turn to formalizing aspects relating to measurement and evaluation.

Definition 8 (Returns)**.** Environment $\mathcal{E}$, population $\mathbf{P}$, and infrastructure $\mathbb{I}$ which produces $\mathbf{P}'$ after some infrastructure response, yields the *infrastructure returns* over all classes K:

$$\underbrace{\sum_{(n'_i, c'_i, \pi'_i) \in \mathbf{P}'} n'_i \cdot \Big(E[\pi'_i | M'_i]\Big) \; - \sum_{(n_i, c_i, \pi_i) \in \mathbf{P}} n_i \cdot \Big(E[\pi_i | M_i]\Big)}_{\text{Change in Agent Costs}} \; + \sum_{(h^{\mathrm{T}}_j, h^{\mathrm{C}}_j, c_j) \in \mathbb{I}} c_j,$$

The expression can be understood as follows: the original MDPs for population $\mathbf{P}$ have an expected cumulative cost over their policies; after infrastructure is applied, the new population $\mathbf{P}'$ may adopt different policies, potentially resulting in a change in costs. The difference between these two values gives the change in expected costs under the infrastructure response, while the final term includes the infrastructure's upfront construction expense. Taken together, the result is the final expected cost incurred by all agents under the response to infrastructure $\mathbb{I}$ on $\mathcal{E}$. This expected cost can be interpreted not only as an estimate for a single agent, but as the average cost over many independent repetitions.[5]

Returns permit determination of "break-even" points for a proposed piece of infrastructure: by changing the original population sizes n_k for each $(n_k, c_k, \pi_k) \in \mathbf{P}$ or by modifying adoption rate α, the change in agent costs may be adjusted until the returns are equal to zero, at which point the original construction costs have been recouped. (See, also, Section V.) While the break-even point for proposed infrastructure depends on the population of each agent class, these populations may also be thought of as a usage rate where each agent of class k must make use of the infrastructure n_k times before saved execution costs equal expended construction costs. Interpretation of the model in such a way disregards potential outside interference from other agents—which is reasonable when the base MDP

[3]Specific instances of $\mathbf{S}(\cdot)$ might be Value- or Policy-Iteration solvers, or some Reinforcement Learning method.

[4]In cases where agents have choice in their initial configuration (cf. footnote 1), an adaptive utilization may result in a different choice in s_0 than the original policy. Naturally, oblivious utilization retains initial state s_0.

[5]Assuming non-interference between agents, the expected cost for a single agent can be easily extended to a group solely through n_i and n'_i terms.

can be assumed to reflect the agents' normal interactions in the environment, already including *some* interference. Section IV's case studies also omit inter-agent interference.

C. Formal Definitions of Informal Concepts of Infrastructure

We return discuss some features of infrastructure treated informally before. The following defines impactfulness:

Definition 9 (Impactful Infrastructure)**.** Given environment $\mathcal{E}'$ derived from $\mathcal{E}$ and infrastructure $\mathbb{I}$, $\mathbb{I}$ is *impactful* if:

1) For at least one adaptive agent class k, $\mathbf{S}(M'_k) \neq \mathbf{S}(M_k)$ or, alternately, $\pi'_k \neq \pi_k$, (under operator $\mathbf{S}(\cdot)$) and
2) $E[\pi'_k|M'_k] \neq E[\pi_k|M_k]$.

This definition reflects the two aspects of "Impacts Agent Behavior" of Section I-A: $\mathbb{I}$ should both alter the operation of agents (through adaptive agents generating a policy different to their original) and have a measurable impact (through the new policy having a different expected cumulative cost).

Impactful infrastructure may further be classified by the manner in which it affects agents, and changes in the expected cost of policies can be used to determine if infrastructure is beneficial to an agent or not. Generally, if the expected cost once an agent has adapted is higher than the cost of the original policy on the original MDP, $E[\pi'|M'] > E[\pi|M]$, the infrastructure is *harmful* to the agent. Conversely, if $E[\pi'|M'] < E[\pi|M]$, the infrastructure has reduced overall expected costs and is considered *beneficial*.

Some infrastructure may also have the goal of inducing behavior change, oftentimes for social engineering aims [25]. Instances that seek to "encourage" behavior π' over π may introduce infrastructure that changes $E[\pi'|M] \geq E[\pi|M]$ to $E[\pi'|M'] \ll E[\pi|M']$. As this entails neither $E[\pi'|M'] < E[\pi|M]$ nor $E[\pi'|M'] > E[\pi|M]$, it is neither directly beneficial nor harmful. Possible modifications could be via a penalty/punishment system, where $E[\pi|M'] \gg E[\pi|M]$ becomes $E[\pi|M'] > E[\pi'|M']$, or via rewards, where $E[\pi'|M] \geq E[\pi|M] \approx E[\pi|M'] \gg E[\pi'|M']$.

For additional formalizations of some properties of infrastructure within the model, we refer the reader to [5, Ch. 4].

IV. Case Studies[6]

Definition 5 builds upon two functions: one modifying transitions (h_i^{T}) and another modifying costs (h_i^{C}). However, the way in which infrastructure can alter agent behavior can vary widely. Broadly, we consider the effects on two dimensions *perception* vs *actuation*, and *precision* vs *efficiency*. The case studies to follow exemplify different ways in which agents can be affected (see Table I).

	Precision	**Efficiency**
Perception	Improved sensing through changes in the environment. Ex: IV-B	Improved efficiency through modifying what is sensed. Ex: IV-D
Actuation	Improved actuation outcomes through changes in the environment. Ex: IV-A	Improved efficiency through modifying actuation outcomes. Ex: IV-C

TABLE I: A guide to the examples presented in Section IV.

[6]All examples in Section IV were implemented in Python and based upon the code of [26]. Policies were found through Value Iteration.

A. Carpeted Care Facility: Increased Actuation Precision

Increasingly, robots are being introduced into long-term care facilities to help handle tasks like medication delivery, laundry, and night patrols [27]. Also, "social" robots to encourage interaction and entertain residents are becoming common. It is important to scrutinize how such robots will interact with residents, staff, and the building itself [28].

Figure 1 shows a communal living space based on existing care facilities in [29]. The layout includes private living spaces for residents, communal areas (e.g., dining and common areas), and staff-only spaces (the reception, kitchen, etc.). Robots are being considered for tasks like delivering medication to residents, retrieving laundry, and escorting residents and visitors around the facility.

Low-pile carpet such as berber carpet is common within residential facilities to reduce dust and noise. Though not a major impediment, compared to tile, these carpets increase the energy needed per unit distance. Friction from the carpet can also reduce the accuracy of rotation and other movements. Also, robots travelling the same routes many times in a day will accelerate carpet wear. Adhesive carpet runners are plastic strips that can be applied on top of carpets. In addition to protecting carpets, they provide a smooth surface that is ideal for a robot to traverse; placed at a hallway edge, they provide an efficient "lane" for robots to use.

Other considerations arise from the residents. Transitions between different flooring is a trip hazard, and therefore the laying of paths should balance safety considerations with efficiency requirements of the robots. In attempting this balance, the runners (shown in yellow) in Figure 1 do not cross across hallways or in front of doorways that lead to carpeted rooms. In keeping these areas clear, the tripping danger is reduced while still spanning much of the facility.

In this problem, the MDP's transition costs reflect the time required to move on different types of flooring. Figure 1 shows the difference in transition probabilities for carpet vs the carpet runners. Aside from improving the precision of the robot's motion, the cost for the action is also reduced. A time penalty is incurred for collisions with obstacles or walls. The robot's initial and goal locations are selected uniformly at random from the locations marked with stars in Figure 1, leading to the expected cost of a robot's trip to be $157.86\,\mathrm{s}$. Keeping the same policy on the updated

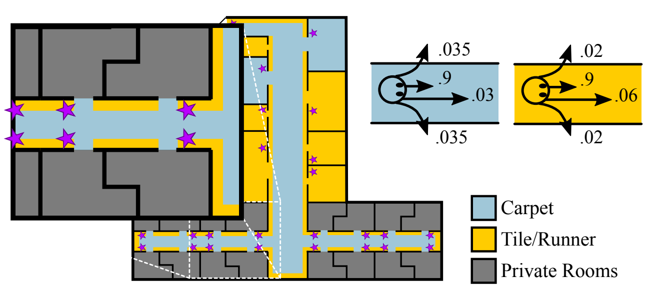

Fig. 1: An example long-term care facility. Purple stars are potential goals. The robot takes $5\,\mathrm{s}$ to travel $1\,\mathrm{m}$ on carpet, and $3\,\mathrm{s}$ to travel $1\,\mathrm{m}$ on tile. Transition probabilities for the floors are shown.

facility, i.e., oblivious use, resulted in a slight reduction (16 % improvement) to an expected trip time of 132.26 s, while adaptive robots had times improved by 36 % to 100.68 s.

In the initial configuration, the robot has no preference for any part of the hallway it travels in, and the resulting policy is just the fastest route. The benefits seen in oblivious use is a result of travelling over runners incidentally. With an updated model, the runners become natural travel routes for the robot. Improved motion dynamics yield policies where robots approach the nearest runner en route their destination—the robots staying near walls has the unintended benefit of reducing interference with residents/staff in the hallways.

1) Conceptual Extensions: A benefit of specific paths for robots is that they could have distance markers or QR codes applied to them. While the current robots in the facility cannot sense this information, future robots might use such patterns for more accurate navigation. Our next case study considers the idea of infrastructure modifying sensing.

B. Material Handling: Perception that Improves Precision

Figure 2 depicts a fulfillment warehouse in which robots assist with loading trucks. Each collects goods from pickup location (A, B, or C) and transports them to a drop-off zone, where other agents then sort and finally load. Where goods appear, and where their drop-off location will be, is assumed to be random (i.i.d.). When not active, the robots occupy one of two maintenance bays; when tasked, the robot could be in either bay with equal likelihood.

The robots use low-resolution cameras to determine if an area is open space or contains an obstacle, but they have low accuracy and perceived obstacles may differ in size from reality. Consequently, the robots do not have sufficient certainty that they will avoid trucks when passing between them, forfeiting some efficiency. The top row of Figure 2 shows policies under these conditions: robots favor the middle path to avoid uncertainty-induced risk. The warehouse manager wants robots to take the shortest possible path to improve operational efficiency and balance avoid congestion. However, any environmental modifications must respect her limited budget. She decides to mark the trucks' parking spots with a high-visibility tape, enhancing contrast between trucks and the floor. This improves identification of obstacles, allowing the robots to move between trucks with less risk.

This problem is modeled as multiple MDPs, sequenced together: Starting in a random maintenance bay, the robot is assigned to pick up goods at one of three locations. After achieving this initial goal, that location becomes its new initial state and it is assigned one of three drop-off locations. Finally, the agent begins at a drop-off spot and is assigned one of two maintenance bays to return to. Perception improvements are modeled through the controller: the initial movement model for agents near obstacles has a chance that the agent will drift when moving forward representing the probability of the low-level control loop driving the system forward when the space ahead is misidentified as free when it is not. More accurate sensing reduces the chance of drift, decreasing expected penalties.

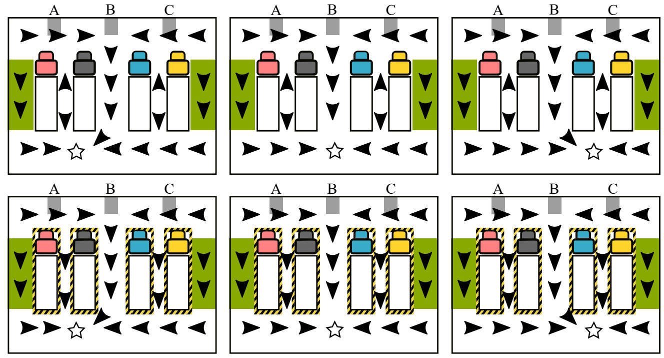

Fig. 2: Various policies for different drop-off locations, indicated with a star. **Top:** Uncertainty in sensing causes agents to take a longer path to avoid obstacles. **Bottom:** High-visibility markings allow agents to pass between trucks safely.

The top and bottom rows of Figure 2 show examples of policies for a robot proceeding from any of the pickup zones to several different drop-off points. As the optimal policy depends on which of the several goals the robot is assigned, it maintains different policies for its various starting locations, pick-up zones, and drop-off locations. The bottom row shows policies after the introduction of high-visibility tape, and deliberate motions between trucks is clearly visible.

The application of high-visibility tape allows for the current agents to remain in use with minimal environmental changes or cost. This change has the largest impact on the cost for agents to transport loads from pickup to drop-off. Given that the regions between trucks are unlikely to be visited during the other parts of the process, the sequencing of MDPs also enables us to uncover precisely where infrastructure offers the greatest benefit (i.e, the transport step).

C. Bridges in the Park: Actuation to Improve Efficiency

Two businesses in nearby buildings—separated by a park but with a roadway connecting them—wish to transport goods back and forth (Figure 3). Both sides maintain a fleet of robots for transporting goods. The park is popular with employees and visitors, who take walks during their breaks. Both robots and people can access the road, though absence of sidewalks means there is some risk of an accident, incurring a high cost. People prefer shorter routes to maximize the area they can visit in the park during their breaks. Robots also prefer a shorter route to reduce travel time as maintenance is performed after a certain number of hours of service.

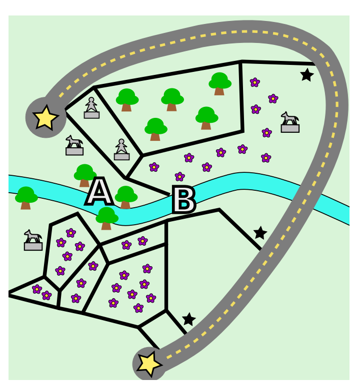

Fig. 3: Park layout highlighting the complexity of choices involved when modifying environments through infrastructure. The pair of large yellow stars indicate the locations of the two businesses, while smaller stars indicate additional entry and exit points for pedestrians. The two locations labeled **A** and **B** are potential sites for a proposed bridge.

The businesses consider two possibilities: (i) speed bumps to slow road traffic and reduce accidents; (ii) a bridge in the park that provides a safer and faster path. While the businesses may favor a bridge positioned to enable fast routes for their robots, employees petition for a bridge facilitating easy travel between landmarks of interest. But what is the benefit of a bridge—for the robot fleet and employees? Will safety and speed justify the expense of bridge construction? If so, at which location should it be built?

To obtain answers, robots will naturally be treated as if solving an MDP. While humans cannot be controlled *per se*, a policy is still a serviceable modeled: we employ a basic MDP constructed via some simple assumptions—a more sophisticated one, based on observations of how they move through the park (say using inverse reinforcement learning [30]) could be used if desired. We indicate the MDP which describes the robots as M_{r}, and the MDP based on human observation M_{h}. The robots may begin at one of the two buildings and have the other as their goal, while the human policy generally has employees returning to the entrance they started at. For this example, we find the optimal policy for the robots, denoted π_{r}. Similarly, we designate the MDPs that have undergone a transformation from infrastructure as M'_{r} and M'_{h}.

The two proposed bridge locations are shown as **A** and **B** in Figure 3. The original MDPs (M_{r} and M_{h}) reflect a world without bridges. Three different infrastructure transformations were performed: one that places a bridge at **A**, one putting a bridge at **B**, and one with bridges at both **A** and **B**. Speed bumps are constructed in all cases. Table II shows expected travel time for various routes, created from the randomly chosen of sub-goals.

Neither robots nor humans take advantage of the infrastructure obliviously as routes are planned to cross bridges only when they're known, while the speed bumps will impact the agents regardless. Humans are unaffected by speed bumps, but the robots have difficulty traversing them and now incur a small additional time cost on the roads.

The adaptation of agents results in a new policy π'_{r}, with an associated expected reward $E[\pi_{\mathrm{r}}|M'_{\mathrm{r}}]$. As M_{r}'s cost function represents the total travel time, the difference between the updated expected cost $E[\pi'_{\mathrm{r}}|M'_{\mathrm{r}}]$ and the original $E[\pi_{\mathrm{r}}|M_{\mathrm{r}}]$ directly reflects the resulting time-saving. For comparison to the implementation cost, the travel time is converted into a dollar amount by pro-rating the cost of maintenance over the amount of time the robot can run between services.

Robot	No Bridge	Bumps	Bridge **A**	Bridge **B**	**A & B**
Route 1	201.7	213.0	115.3	101.9	101.9
Route 2	199.4	210.6	115.3	101.9	101.9
Route 3	236.0	243.7	148.9	119.3	119.3
Route 4	223.5	234.8	211.4	198.0	198.0
Route 5	353.2	361.0	192.0	220.5	192.0
Average	242.8	252.6	156.6	148.3	142.6

Human	No Bridge	Bumps	Bridge **A**	Bridge **B**	**A & B**
Route 1	914.9	914.9	455.1	549.5	455.1
Route 2	672.6	672.6	446.9	348.2	348.2
Route 3	303.3	303.3	303.3	303.3	303.3
Route 4	744.3	744.3	536.2	463.7	463.7
Route 5	1437.9	1437.9	514.1	587.4	451.9
Average	814.6	814.6	451.1	450.4	404.4

TABLE II: Expected travel times in seconds given the two different bridge locations. The *bumps* column marks the introduction of speed bumps, but no use of bridges. **Top:** The robot fleet moves at a speed of $5\,\mathrm{km/h}$. **Bottom:** The humans move at $3\,\mathrm{km/h}$.

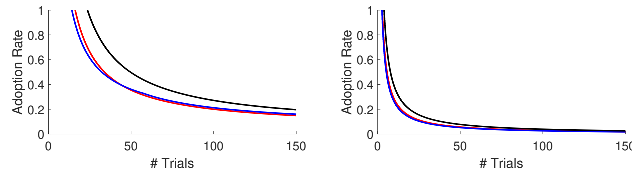

Fig. 4: Impact of adoption rate and the number of infrastructure uses on recouped costs. These graphs show, for robots **(left)** and humans **(right)**, the number of uses needed at different adoption rates in order for saved costs to equal construction costs. Red lines correspond to Bridge **A**, blue lines to Bridge **B**, and black lines to both bridges simultaneously. Each line is a different route.

If robots do not adapt to the new environment, the additional time incurred by the speed bumps results in an increase in overall travel time. The introduced infrastructure therefore is detrimental to oblivious robot agents. However, once either bridge is introduced and the robots adapt and change their policies, the robot agents' costs are reduced.

Figure 4 shows, for both types of agents, the number of trips the agents must take to recoup the infrastructure's construction costs. Using the equation in Section III, there exists a non-linear relation between the adoption rate α and the number of trips necessary. For the robots, who incur additional costs on their original routes, higher adoption rates are necessary to offset costs of the oblivious part of the fleet.

We now consider the impact of infrastructure on the people. The assignment of a monetary cost to their trips through the park is more difficult than for the robots. Although the humans are not looking to enter and leave the park as fast as possible (unlike the robots), there are other factors (such as the duration of their break) that mean they still benefit from bridges that reduce their route and prevent them from having to use the roadways. For clarity in this example (and, perhaps, somewhat bleakly), we will consider the time spent on paths to correspond to time spent not working, and therefore shorter paths result in increased profits for the business. Additionally, the introduction of bridges results in humans avoiding the road. This results in far fewer accidents, which is a strong indication that a bridge is worth its construction cost.

The final selection of where to place the bridge gives rise to a consideration of "fairness," wherein the businesses must compare the impact of the bridges on both robots and humans. Suppose that Bridge **A** costs $\$1600$, and Bridge **B** costs $\$1200$. The cost of the infrastructure is recouped directly through reduced traffic accidents and maintenance costs, and indirectly recouped through employees. As employees are not negatively impacted by the introduction of speed bumps on the road, their graph does not show much variation between bridge options; Bridge **B** results in slightly shorter paths on average for employees than Bridge **A**, but both show that costs are quickly recouped even when

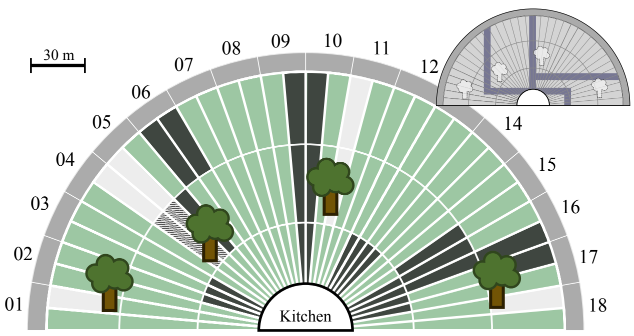

Fig. 5: An indoor atrium surrounded by guest rooms. It is divided up into states based on 5° segments of three concentric rings. Sources of sensing interference include trees and underground pipes (see inset). Legend: *White:* with a beacon located at the kitchen, trees within the atrium cause interference when line-of-sight is broken. *Black:* with a fluxgate compass, underground pipes cause interference. *Striped:* interference for both sensors.

adoption rates are low. For the robot fleet, the introduction of speed bumps and resulting additional costs requires agents to adapt before any benefit is seen. The trending of the graphs towards a limit as adoption rate decreases also suggests that for a given expected amount of usage there is a minimum number of agents who must adapt for the bridge to be practical. Overall, Bridge **B** recoups costs more quickly than Bridge **A**. Given this, and the lack of strong employee preference, Bridge **B** would be the best option.

D. The Hotel: Perception to Improve Performance

Figure 5 shows an example of a hotel with an indoor atrium. Bordering the atrium are 18 rooms for guests, as well as a kitchen for room service. The kitchen uses robots to deliver room service orders (this is one task the service robots may soon play in the hospitality industry, see [31] and references therein). The robot makes use of a fluxgate-like compass to help determine its position and heading in the hotel, and occasionally pauses at certain waypoints to let its sensor settle and obtain an updated reading.

Each room can be set up for four different occupants: a family, an individual on vacation, a couple, or a business traveler. Different occupants have different values for the daily average number of orders they make to room service. Rooms can also be unoccupied, in which case no orders are generated from them. The hotel has two questions: first, given several different potential room layouts, which results in the least travel time for the robots (and thus greater efficiency)? Second, the hotel is considering upgrading the sensors on the robots to a beacon system (such as [32]); what is the benefit of this upgrade, considering the new layouts?

For this example, the cost function includes both the time it takes for the robot to complete its movement as well as the time taken for the robot to obtain and verify a sensor reading. When the robot ends up in a state other than the one intended, there is an additional delay as it confirms this with another sensor reading. The atrium itself is divided into three concentric rings, each divided into 5° slices. These represent the way the robot periodically checks its sensors to ensure that it is staying on course. We use a Poisson distribution

	Average Travel Time (s)		
	No Interference	Compass	Beacon
Layout 1	2934	2970	2942
Layout 2	2997	3033	3006
Layout 3	2964	2999	2975

TABLE III: Comparison of average time spent travelling each hour given an average of 10 meal orders per hour, three different guest layouts and robot speed of $5\,\mathrm{km/h}$. Values given are for the model under three interference models: no interference, interference to a beacon, and interference to a compass (Fig. 5).

($\lambda = 10$) to model the number of orders received by the kitchen each hour. For each order, the robot's goal is selected with a weighting based on the allocation of guest rooms.

Both types of sensor (compass and beacon) have areas in which they receive interference. The compass receives interference from underground pipes, while there exist certain blind spots from where the robot cannot see the beacon, and must estimate its position through dead reckoning (see Figure 5). In both cases, these cause the robot to take slightly longer to obtain a reading for its position.

Table III shows three potential room layouts where all rooms are in use: (1) one where rooms are assigned at random, (2) one where groups are housed closer to the kitchen to concentrate noise, and (3) one where rooms are allocated in a repeating pattern. We can compare the average travel times under no interference, interference imposed upon a compass, and interference imposed upon a beacon.

Over all three layouts, the robot with a beacon performs better than the robot with a compass. This result is easily explained by looking at Figure 5, where we see that the compass has 27 regions where underground pipes cause interference, as opposed to the 8 regions where the beacon receives interference. However, Table III also shows that despite this large difference, the actual impact of this interference is quite small: the robot equipped with the fluxgate compass spends on average just over $26\,\mathrm{s}$ longer delivering meals each hour. Considering that this difference would be distributed over multiple orders in an hour, it becomes almost negligible when considered against other factors, such as the time it takes for hotel guests to take their food when it arrives.

V. Maintenance

Infrastructure can be costly to maintain. Worn-out carpet runners must be replaced, bridges repaired, footpaths repaved, and so on. Including maintenance costs into the equations given in Section III requires first that the reduced costs seen by the agents be matched to a timescale. We will revisit the example from Section IV-A to demonstrate the general steps needed. The cost of carpet runners of the type described is dependent on thickness and width, but for this example we will assume they cost $\$10/\mathrm{m}$. Runners are replaced once a year to remove any that have become discolored or cracked. The facility requires just over $150\,\mathrm{m}^2$ of runners, leading to a yearly maintenance cost of $\$1500$.

Any savings the robots induce must also be estimated over a year, a highly problem-dependent calculation. We assume that the facility has three robots and, at a minimum, these robots make 48 trips per day, delivering medication three times a day to each resident. They likely carry laundry back and forth, and guide residents around but, for a pessimistic estimate of break-even point, we will ignore these uses.

Without the runners, the expected time of a delivery trip was $157.86\,\mathrm{s}$. With the runners, this decreased to $100.68\,\mathrm{s}$. This results in a daily average of $7577.28\,\mathrm{s}$ and $4832.64\,\mathrm{s}$, respectively. With the runners and this minimum amount of trips, the robots reduce their time spent traveling daily by just under $46\,\mathrm{min}$. To assign a monetary value to this, assume each robot frees up a nursing assistant to perform other duties. Therefore, valuing the work of the robots at the cost of employing a human for the same amount of time at $\$16/\mathrm{h}$ the "cost" of the robots becomes $\$33.68$ per day and $\$21.49$ per day. Given that care facilities operate every day, the yearly maintenance cost will be recouped after 123 days.

VI. Conclusion and Future Work

This work connects environmental modification with robot design by introducing a model for analysis of the wide variety of forms infrastructure takes. The model treats infrastructure as an operator that transforms existing MDP models to reflect changes to behavior induced by altering the environment. Within the framework, we interpret the MDP's expected value not as a representative statistic for a single agent making decisions about an uncertain future, but as the cost over many independent repetitions. This allows us to understand the average behavior of a class of agent without direct large-scale simulation of all agents involved.

The model has room for refinement, which future work could explore. As noted, the model disregards inter-agent interference—this was a deliberate choice to allow tractable analysis of aggregate effects across many repetitions. In certain settings, contrary to the case studies examined herein, interference may be critical and models of how interactions manifest would be a useful addition. One simple option for obtaining an effective model might be the inclusion of a correction term (e.g., the penalization function $Q(\cdot)$ of [33]).

Infrastructure itself can also be more complex, containing internal state. The need to capture complex behavior and the impact of infrastructure as something that interferes with agents (via its transformation functions) suggests that infrastructure itself may be modeled as a type of agent.

References

[1] A. Censi, "A Class of Co-Design Problems With Cyclic Constraints and Their Solution," *IEEE Robotics and Automation Letters*, vol. 2, no. 1, pp. 96–103, Jan. 2017.

[2] L. Carlone and C. Pinciroli, "Robot Co-design: Beyond the Monotone Case," in *Proc. IEEE International Conf. on Robotics and Automation*, Montreal, Canada, May 2019.

[3] R. Desai, J. McCann, and S. Coros, "Assembly-aware design of printable electromechanical devices," in *Proc. ACM Symposium on User Interface Software and Technology (UIST)*, 2018, pp. 457–472.

[4] D. A. Shell, J. M. O'Kane, and F. Z. Saberifar, "On the design of minimal robots that can solve planning problems," *IEEE Trans. on Automation Sci. and Eng.*, vol. 18, no. 3, pp. 876–887, 2021.

[5] G. A. McFassel, "Sensing and Infrastructure Design for Robots: A Plan-Based Perspective," PhD Dissertation, Texas A&M University, Department of Computer Science & Engineering, Aug. 2023.

[6] G. Zardini, N. Lanzetti, A. Censi, E. Frazzoli, and M. Pavone, "Co-Design to Enable User-Friendly Tools to Assess the Impact of Future Mobility Solutions," arXiv 10.48550/arXiv.2008.08975, 2020.

[7] E. Bonabeau, M. Dorigo, and G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press, Inc., 1999.

[8] R. Beckers, O. E. Holland, and J.-L. Deneubourg, "From Local Actions to Global Tasks: Stigmergy and Collective Robotics," in *Artificial Life IV*, Cambridge, MA, U.S.A., Jul. 1994, pp. 181–189.

[9] A. J. Ijspeert, A. Martinoli, A. Billard, and L. M. Gambardella, "Collaboration through the exploitation of local interactions in autonomous collective robotics: The stick pulling experiment," *Autonomous Robots*, vol. 11, no. 2, pp. 149–171, 2001.

[10] D. Payton, M. Daily, R. Estowski, M. Howard, and C. Lee, "Pheromone Robotics," *Autonomous Robots*, vol. 11, no. 3, pp. 319–324, Nov. 2001.

[11] R. T. Vaughan, K. Støy, M. J. Matarić, and G. S. Sukhatme, "LOST: Localization-Space Trails for Robot Teams," *IEEE Trans. on Robotics and Automation*, vol. 18, no. 5, pp. 796–812, Oct. 2002.

[12] B. T. Fine and D. A. Shell, "Eliciting collective behaviors through automatically generated environments," in *Proc. IEEE/RSJ International Conf. on Intelligent Robots and Systems (IROS)*, 2013, pp. 3303–3308.

[13] L. Bobadilla, O. Sanchez, J. Czarnowski, K. Gossman, and S. M. LaValle, "Controlling wild bodies using linear temporal logic," in *Robotics: Science and systems*, vol. 7, 2012, p. 17.

[14] T. Sandler and J. T. Tschirhart, "The Economic Theory of Clubs: An Evaluative Survey," *Journal of Economic Literature*, vol. 18, no. 4, pp. 1481–1521, Dec 1980.

[15] J. M. Buchanan, "An Economic Theory of Clubs," *Economica*, vol. 32, no. 125, pp. 1–14, Feb 1965.

[16] G. Owen, *Game Theory*. New York: Academic, 1982.

[17] S. LaValle and S. Hutchinson, "Game theory as a unifying structure for a variety of robot tasks," in *Proc. IEEE International Symposium on Intelligent Control*, 1993, pp. 429–434.

[18] T. Roughgarden, "Algorithmic game theory," *Communications of the ACM*, vol. 53, no. 7, pp. 78–86, 2010.

[19] M. G. McNally, "The Four Step Model," in *Handbook of Transport Modeling, Second Edition*. Amsterdam ; Oxford: Elsevier, 2007, ch. 3.

[20] G. O. Kagho, M. Balac, and K. W. Axhausen, "Agent-based models in transport planning: Current state, issues, and expectations," *Procedia Computer Science*, vol. 170, pp. 726–732, April 2020.

[21] T. Krendl Gilbert, A. J. Snoswell, M. Dennis, R. McAllister, and C. Wu, "Sociotechnical Specification for the Broader Impacts of Autonomous Vehicles," 2022, in *IEEE ICRA Workshop on Fresh Perspectives on the Future of Autonomous Driving*, available via arXiv 10.48550/arXiv.2205.07395.

[22] H. Mirzaei, G. Sharon, S. Boyles, T. Givargis, and P. Stone, "Enhanced delta-tolling: Traffic optimization via policy gradient reinforcement learning," in *Proc. International Conf. on Intelligent Transportation Systems (ITSC)*, nov 2018, pp. 47–52.

[23] F. Rasheed, K.-L. A. Yau, R. M. Noor, C. Wu, and Y.-C. Low, "Deep reinforcement learning for traffic signal control: A review," *IEEE Access*, vol. 8, pp. 208 016–208 044, 2020.

[24] D. P. Bertsekas, *Reinforcement Learning and Optimal Control*. Belmont, M.A., U.S.A: Athena Scientific, 2019.

[25] R. H. Thaler and C. R. Sunstein, *Nudge: Improving decisions about health, wealth, and happiness*. Penguin, 2009.

[26] S. J. Russell and P. Norvig, "Python implementation of algorithms from "Artificial Intelligence—A Modern Approach"." [Online]. Available: https://github.com/aimacode/aima-python

[27] T. Jacobs and B. Graf, "Practical evaluation of service robots for support and routine tasks in an elderly care facility," in *IEEE Workshop on Advanced Robotics and its Social Impacts*, 2012, pp. 46–49.

[28] M. Niemelä and H. Melkas, *Robots as Social and Physical Assistants in Elderly Care*. Singapore: Springer Singapore, 2019, pp. 177–197.

[29] W. Benbow. (2011, January) Dementia Design: H Shape Facility Layout. Accessed 02-26-2022. [Online]. Available: https://wabenbow.com/?page_id=252

[30] S. Adams, T. Cody, and P. A. Beling, "A survey of inverse reinforcement learning," *A.I. Review*, vol. 55, no. 6, pp. 4307–4346, 2022.

[31] I. Y. Lin and A. S. Mattila, "The Value of Service Robots from the Hotel Guest's Perspective: A Mixed-Method Approach," *International Journal of Hospitality Management*, vol. 94, p. 102876, 2021.

[32] J.-S. Gutmann, P. Fong, L. Chiu, and M. E. Munich, "Challenges of designing a low-cost indoor localization system using active beacons," in *IEEE Conf. on Technologies for Practical Robot Applications*, 2013.

[33] C. Nam and D. A. Shell, "Assignment Algorithms for Modeling Resource Contention in Multi-Robot Task Allocation," *IEEE Trans. on Automation Sci. and Eng.*, vol. 12, no. 3, pp. 889–900, Jul. 2015.