Measuring Target Predictability for Optimal Environment Design

Melkior Ornik

***Abstract*— Motivated by the study of deceptive strategies, this paper considers the problems of detecting an agent's objective from its partial path and determining an optimal environment to enable such detection. We focus on a scenario where the agent's objective is to reach a particular target state from a set of potential targets, while an observer seeks to correctly identify such a state prior to the agent reaching it. In order to quantify the predictability of the agent's target given the observed path, we introduce the notion of target entropy, where higher entropy implies lower target predictability. The problem of optimal environment design, i.e., optimal target placement, then becomes a minimax problem with target entropy as an objective function. Under the assumption that the agent chooses its path towards its target maximally unpredictably, we consider models of the agent's motion on both discrete and continuous state spaces. Using dynamic programming, we establish a simple way of computing target entropy for the discrete state space. In a continuous state space, we obtain a formula for target entropy by employing geometrical arguments on volumes of hypersimplices. Additionally, we provide an algorithm yielding an optimal environment in a discrete state space, discuss its computational complexity, and provide a computationally simpler approximation that yields a locally optimal environment. We validate our results on a previously developed model of deceptive agent motion.**

I. Introduction

The problem of deceptive motion and its detection has been in focus of recent research, both in terms of mathematical treatment — models, optimal deceptive strategies, and equilibria [2], [12], [13], [17] — and in empirical studies on humans [1], [6], [11]. Such work is based on the following premise: lacking any communication with an agent or other information, the only feature that can be used to determine an agent's objective is its path. A focus of research in deception is, thus, on planning a path that is most likely to instill incorrect information about the location of its target to an observer for the longest possible time while still arriving at the desired target [6], [14], [19].

This paper deals with a natural dual to deceptive path planning: *goal recognition*, i.e., inference of the agent's objective from its path during some initial observation period. Previous discussions on such a problem [12], [15], [16], [17] were often solely presented through the lense of path planning: inference was conditioned on assuming some specific knowledge about the agent's strategy, and assuming that the observer has no influence on the environment, e.g., on the agent's dynamics or locations of the agent's possible targets. The former assumption is difficult to support when dealing with an unpredictable, previously unseen, and potentially deceptive adversary. The latter assumption is incorrect in a wide variety of scenarios where the observing party has control over the environment: for instance, facility planners naturally position features in the environment in such a way to enable identification of visitors and their intentions [20], [21]. Consequently, this paper is the first step at establishing *optimal environment design* as a fundamental formal problem in the theory of counterdeception and goal recognition. Fig. 1 gives an example of two environments differing in their receptiveness to deception.

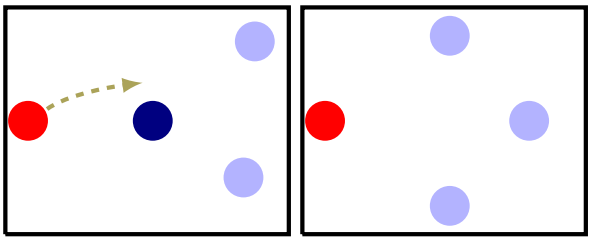

Fig. 1. Two environments, i.e., target placements, that are not equally receptive to deception. An agent starting from the red point in the left environment may believably appear to be moving towards either the top or bottom target (marked by light blue), for instance by initially following the yellow path, while actually proceeding towards the middle target (dark blue). The intentions of an agent in the right environment necessarily become obvious well before reaching the intended target. We note that the counterdeceptive quality of the environment does not depend solely on the distance between potential targets.

The problem of environment design with the intent of assisting in identifying an agent's target is related to previous efforts on facility planning [7], [10]. Such work seeks to find an optimal placement of sites within a given area according to a particular criterion, e.g., distance to a particularly undesirable site [7]. However, to the best of our knowledge, previous work on location planning does not consider our purpose — identifying an agent's objective — and thus yields substantially different metrics of optimality and optimization problems from our paper.

This paper primarily considers the scenario of a multi-site facility where the attacker is known to want to enter exactly one of the sites, but it is not known which one. The environment design, which we in this paper limit to placement of the sites, is left to the defender, who subsequent attempts to identify the attacker's true target as the attacker is moving. While the scenario of multiple possible targets has been extensively used in previous work on deception [12], [13], [17], the problem of optimal target placement has, to the best of our knowledge, not been previously considered.

Development of a meaningful metric of the ability of a particular target placement to aid in identifying the attacker's

M. Ornik is with the Department of Aerospace Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign. Email: mornik@illinois.edu

target is a fundamental challenge to optimal target placement. While such a metric could be derived from specific previously developed models of the agent's behavior and dynamics of observer's beliefs about targets [17], such an approach has two pitfalls: (i) it may be computationally difficult, depending on the complexity of these dynamics, and (ii) it would only provide a metric that corresponds to those particular behaviors, and thus could not yield a target-revealing environment maximally agnostic of the agent's possible behavior. Correspondingly, in this paper we explicitly make the assumption that the agent chooses its target and path as unpredictably as possible, both with a uniform probability distribution.

The primary contribution of the paper is to define a metric of *target entropy* of an environment at a given time, as the maximal entropy of the *target prediction vector* over the state space. We calculate the target prediction vector from the proportion of paths that pass through the given state at a given time while continuing to their target; consequently, we provide a method for easy computation of the target prediction vector both in a discrete and a continuous state space. We subsequently consider the problem of optimizing the target placement to minimize target unpredictability. Limiting the remainder of the work to the discrete state space, we discuss the computational complexity of a naive optimization algorithm and provide a computationally lighter algorithm that finds a locally optimal environment.

We validate the obtained optimal environments by measuring the success of a observer at identifying the agent's target using the optimal deception strategy and defender belief model developed in [17]. Our method yields significantly higher probability of target identification, and thus significantly worse results for a deceptive agent, than the environment considered in [17].

Notation: For any $n \in \mathbb{N}$, $[n]$ denotes the set $\{1, 2, \dots, n\}$. Given a set S, $\#(S) \in \mathbb{N} \cup \{0\} \cup \{\infty\}$ denotes its cardinality, while $|S|$ denotes its volume according to an underlying probability measure. For a real-valued random variable X on an underlying probability space, $\mathbb{E}[X]$ denotes its expected value, and for events A and B on an underlying probability space, $\mathbb{P}(A)$ denotes the probability of the occurrence of A, while $\mathbb{P}(A|B)$ denotes the probability of the occurrence of A conditioned on B; unless stated otherwise, we take by convention that $\mathbb{P}(B) = 0$ implies $\mathbb{P}(A|B) = 0$. If $x \in \mathbb{R}^n$ is a vector, its coordinates are denoted by x_i, $i \in [n]$. Symbol $\log$ denotes the natural logarithm.

II. Problem Statement

We consider an agent operating on the Euclidean space $\mathbb{R}^n$ in discrete time, with dynamics

$$x(t+1) = x(t) + u(t), \quad x(0) = 0, \tag{1}$$

where $u(t) \in \mathcal{U}(x(t))$, and with initial state $x(0) = 0$. We consider two types of allowed sets of actions $\mathcal{U}$: in Section III we allow $\mathcal{U}(x)$ to be any finite set, effectively resulting in a discrete state space, while in Section IV we consider a continuous set of actions $\mathcal{U}(x) = [-1, 1]^n$ for all $x \in \mathbb{R}^n$. While dynamics (1) are simple, they have been extensively used in related work on inference and deception [13], [17]. We say that a T'-tuple $(x(0), \dots, x(T')) \subseteq \mathbb{R}^{n \times (T'+1)}$ is a *path* of length T' if $x(0) = 0$ and $x(t) - x(t-1) \in \mathcal{U}(x(t-1))$ for all $t \in [T']$.

We assume that the agent's objective is to visit a particular *target* state R exactly at time T, i.e., $x(T) = R$. State R is contained in the set of N *potential targets* $\mathcal{R} = \{R_1, \dots, R_N\} \subseteq \mathbb{R}^n$, where we assume $R_i \neq R_j$ for $i \neq j$. The agent's state at any time is known to the observer, who also knows the set $\mathcal{R} = \{R_1, \dots, R_N\}$, but does not know which of the potential targets R_i is the agent's true target R. We consider two problems:

(i) quantifying an observer's uncertainty about the true target R, given the observed partial agent's path $x^{\leq K} = (x(0), \dots, x(K))$ and
(ii) prior to the beginning of the agent's motion, placement of the set $\mathcal{R}$ in such a way that the observer's uncertainty is minimal.

Throughout the paper, we refer to the placement of the set $\mathcal{R}$ as *environment design*. In order to formalize the above problems, we make the following assumptions:

Assumption 1: The agent is known to choose its target R from set $\mathcal{R}$ with uniform probability.

Assumption 2: After choosing its target R, the agent is then known to choose its path to R with uniform probability from the set of all paths of length T that satisfy $x(T) = R$.

Assumption 1 and Assumption 2 can be interpreted as a lack of bias in terms of the agent behavior; while previous works [13], [17] often describe particular models of a deceptive or trustworthy agent, we are seeking to model the behavior of a maximally unpredictable agent [18]. We note that Assumption 2 differs from assuming that the agent's path is a random walk, where an action at every time is made at random independently of previous actions. Instead, the agent randomly a priori chooses its entire path.

For infinite sets of actions $\mathcal{U}(x)$, Assumption 2 requires us to define a meaningful probability measure on the set of all paths such that $x(T) = R$. We formally describe such a measure for the case of $\mathcal{U}(x) = [-1, 1]^n$ in Section IV. For now, we merely assume that such a measure exists.

Let us now formalize the problem of quantifying an observer's uncertainty about the true target. Let $\phi^{\leq K}$ denote the partial agent's path until time K. Assuming that the observer has no prior bias, its belief about the agent's target, if the observed partial agent's path is $\phi^{\leq K} = x^{\leq K}$, is given by the *target prediction vector*

$$P = \begin{pmatrix} \mathbb{P}(R = R_1 \mid \phi^{\leq K} = x^{\leq K}) \\ \mathbb{P}(R = R_2 \mid \phi^{\leq K} = x^{\leq K}) \\ \vdots \\ \mathbb{P}(R = R_N \mid \phi^{\leq K} = x^{\leq K}) \end{pmatrix}. \tag{2}$$

Since (2) describes a probability distribution of a random variable, a natural measure of uncertainty about the random variable is the Shannon entropy [4]. The problem of quantifying the observer's uncertainty about the agent's target is thus formalized as follows.

Problem 3: Let $\mathcal{R} = \{R_1, \ldots, R_N\} \subseteq \mathbb{R}^n$ and $K, T \in \mathbb{N}$, $K < T$. Let $x^{\leq K} = (x(0), \ldots, x(K))$ be a path of length K. If P is given by (2), compute *target entropy*

$$H(P|x^{\leq K}, \mathcal{R}) = -\sum_{i=1}^{N} P_i \log P_i. \tag{3}$$

In the case of $P_i = 0$, we adopt the usual convention [4] of $P_i \log P_i = 0$.

As previously mentioned, the ultimate purpose of our effort is not in solely computing $H(P|x^{\leq K}, \mathcal{R})$, but in designing the environment, i.e., placing potential targets $\mathcal{R}$, in such a way that the agent's target is most predictable.

Problem 4 (Optimal environment design): Let $S \subseteq \mathbb{R}^n$, and $N, K, T \in \mathbb{N}$, $K < T$. Determine

$$\operatorname*{argmin}_{\mathcal{R}} \max_{x^{\leq K}} H(P|x^{\leq K}, \mathcal{R}),$$

where $\mathcal{R} \subseteq S$, $\#(\mathcal{R}) = N$, and $x^{\leq K}$ is a path of length K.

Solving the former problem is necessary to approach the latter. We thus first present the solution to Problem 3.

III. Finite Sets of Actions

In this section, we consider the case where all sets $\mathcal{U}(x)$, $x \in \mathbb{R}^n$, are finite. In such a scenario, given that $x(0) = 0$, there is a finite number of paths that reach a target in $\mathcal{R}$ at time T. As by Assumption 2 all paths are equally probable, we can thus directly proceed to compute probability $\mathbb{P}(R = R_i \mid \phi^{\leq K} = x^{\leq K})$ from Bayes' theorem. We obtain

$$\begin{aligned}\mathbb{P}(R = R_i \mid \phi^{\leq K} = x^{\leq K}) &= \frac{\mathbb{P}(x(T) = R_i, \phi^{\leq K} = x^{\leq K})}{\mathbb{P}(\phi^{\leq K} = x^{\leq K})} \\ &= \frac{\mathbb{P}(x(T) = R_i, \phi^{\leq K} = x^{\leq K})}{\sum_{j=1}^{N} \mathbb{P}(x(T) = R_j, \phi^{\leq K} = x^{\leq K})}.\end{aligned} \tag{4}$$

For any j, $\mathbb{P}(x(T) = R_j, \phi^{\leq K} = x^{\leq K})$ can be computed using the following lemma.

Lemma 5: Probability $\mathbb{P}(x(T) = R_j, \phi^{\leq K} = x^{\leq K})$ equals $\frac{v_j}{N w_j}$, where v_j is the number of paths to R_j of length T that have $x^{\leq K}$ as their initial component and w_j the number of all paths to R_j of length T.

Proof: Quantity $\mathbb{P}(x(T) = R_j, \phi^{\leq K} = x^{\leq K})$ can be interpreted by as the probability that the agent first chooses its target R_j, and then happens to choose one of the paths to R_j of length T that has $x^{\leq K}$ as its initial component. By Assumption 1, the former probability is $1/N$. By Assumption 2, the latter probability equals v_j/w_j. ∎

The value of v_j from Lemma 5 clearly equals the number of paths from $x(K)$ to R_j in time $T - K$. Both v_j and w_j can thus be found from the following result.

Lemma 6: Let $y_S, y_E \in \mathbb{R}^n$ and $T' \in \mathbb{N}$. The number of paths from y_S to y_E of length T' equals $p^{T'}(y_S; y_E)$, where p satisfies the following recursion:

$$p^t(y'; y_E) = \sum_{u \in \mathcal{U}(y')} p^{t-1}(y' + u; y_E), \tag{5}$$

with $p^0(y_E; y_E) = 1$ and $p^0(y'; y_E) = 0$ for all $y' \notin Y_E$.

Proof: Each path from y' to y_E of length t corresponds to exactly one path from some $y' + u$, $u \in \mathcal{U}(y')$, to y_E of length $t - 1$. ∎

From (5), $p^t(y_S; y_E)$ can be computed with complexity linear in t, $\max_x \mathcal{U}(x)$, and the number of states from which there could exist a path to y_E of length less than or equal to t, assuming those states are a priori known. For specific sets $\mathcal{U}(x)$, closed formulae for $p^t(y_S; y_E)$ may exist, e.g., [3] gives an expression for $n = 2$ and $\mathcal{U}(x) = \{e_1, e_2, -e_1, -e_2\}$. In any case, by combining Lemma 5 and Lemma 6, we obtain the following result.

Theorem 7: Let $\mathcal{R} = \{R_1, \ldots, R_N\} \subseteq \mathbb{R}^n$ and let $x^{\leq K} = (x(0), \ldots, x(K))$ be a path of length K. Assume that there exists a path ϕ of length T that reaches $x(K)$ at time K and reaches some $R_{i'} \in \mathcal{R}$ at time T. Then, $H(P|x^{\leq K}, \mathcal{R})$ is given by $H(P|x^{\leq K}, \mathcal{R}) = -\sum_{i=1}^{N} P_i \log P_i$, where

$$P_i = \frac{p^{T-K}(x(K); R_i)/p^T(0; R_i)}{\sum_{j=1}^{N} p^{T-K}(x(K); R_j)/p^T(0; R_j)}, \tag{6}$$

and p is defined as in (5).

Proof: The claim follows directly from the definition of $H(P|x^{\leq K}, \mathcal{R})$ in (3) and (2), expression for $\mathbb{P}(R = R_i \mid \phi^{\leq K} = x^{\leq K})$ in (4), Lemma 5, and Lemma 6. ∎

Naturally, if there does not exist a path of length T to a state in $\mathcal{R}$ that passes through $x(K)$ at time K, the notion of entropy is meaningless: all probabilities P_i equal 0. We now proceed with the discussion of target unpredictability for an agent operating on an n-dimensional state space with continuous actions.

IV. Continuous State Space

In this section, we consider the case of $\mathcal{U}(x) = [-1, 1]^n$ for all $x \in \mathbb{R}^n$; the agent can thus eventually reach any state in $\mathbb{R}^n$. We use the approach presented in Section III to enumerate the paths from $x(0) = 0$ to a target R_i, but draw from measure theory to establish the relevant probabilities.

As mentioned when discussing Assumption 2, computing the probability $P_i = \mathbb{P}(R = R_i \mid \phi^{\leq K} = x^{\leq K})$ is only meaningful if one can define a uniform probability density function on the set of all paths of length T that satisfy $x(T) \in \mathcal{R}$. Since $(x(0), \ldots, x(T)) \in \{0\} \times \mathbb{R}^{n \times T}$, we will show that, if nonempty, the set of all paths such that $x(T) \in \mathcal{R}$ is of finite volume with respect to the usual Lebesgue measure in some $\mathbb{R}^d$, $d \leq Tn$. This feature will allow us to define a uniform probability density function for Assumption 2.

To more easily discuss the geometry of the problem, in the remainder of the section we identify a path $(x(0), \ldots, x(T))$ with control inputs $(u(0), \ldots, u(T-1)) \in [-1, 1]^{n \times T}$ that generate such a path. Set Φ of all paths of length T that reach $\mathcal{R}$ at time T is then given by $\Phi = \{(u(0), \ldots, u(T-1)) \in [-1, 1]^{n \times T} \mid u(0) + \ldots + u(T-1) \in \mathcal{R}\}$. Hence, $\Phi = \cup_{i=1}^{N} \prod_{j=1}^{n} \mathcal{H}^T(R_{ij})$, where $\mathcal{H}^T(R_{ij}) = \{(u_j(0), \ldots, u_j(T-1)) \in [-1, 1]^T \mid u_j(0) + \ldots + u_j(T-1) = R_{ij}\}$, and R_{ij} is the j-th coordinate of R_i. Set $\mathcal{H}^T(R_{ij})$ is a *hypersimplex* [5], [8]: it is the intersection of the hypercube $[-1, 1]^T$ with the hyperplane $\{s \in$

$\mathbb{R}^T \mid \sum_{j=1}^{T} s_j = R_{ij}\}$. Fig. 2 provides an example of the conversion of a path to an element in the product of hypersimplices.

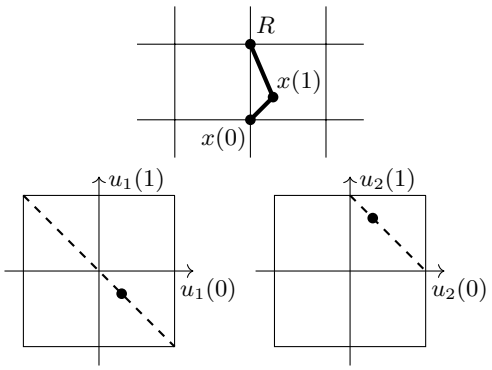

Fig. 2. The top picture shows a path from $x(0)$ to $x(2) = R$. The bottom pictures display u_1 and u_2 that resulted in such a path. Any path from $x(0)$ to $x(2) = R$ such that $u(t) \in [-1, 1]^2$ is uniquely obtained by choosing u_1 and u_2 that lie on the dashed hypersimplices in the bottom two pictures.

By the properties of Lebesgue measures of unions and Cartesian products [9], the set of all paths that reach R_i at time T, interpreted as the set of all control inputs that generate these paths, thus has the finite volume of $V^T(R_i) = \prod_{j=1}^{n} v^T(R_{ij})$, where $v^T(R_{ij})$ is the volume of the hypersimplex $\mathcal{H}^T(R_{ij}) \subseteq \mathbb{R}^T$. For the sake of simplicity, we will only consider the case where either $\mathcal{H}^T(R_{ij})$ is empty or $v^T(R_{ij}) > 0$, while understanding that the $\mathcal{H}^T(R_{ij})$ might be non-empty and have a positive finite volume in some lower-dimensional Lebesgue measure.

To give meaning to Assumption 2, we can now define a probability density function on the set of all paths Φ. To a path, i.e., control inputs $(u(0), \ldots, u(T-1))$, we associate a probability density function

$$f(u(0), \ldots, u(T-1)) = \frac{1}{N \prod_{j=1}^{n} v^T(u_j(0) + \ldots + u_j(T-1))}. \tag{7}$$

We proceed to write $\mathbb{P}(x(T) = R_i \mid \phi^{\leq K} = x^{\leq K})$ as in Section III:

$$\begin{aligned} P_i &= \mathbb{P}(x(T) = R_i \mid \phi^{\leq K} = x^{\leq K}) \\ &= \frac{\mathbb{P}(x(T) = R_i, \phi^{\leq K} = x^{\leq K})}{\mathbb{P}(\phi^{\leq K} = x^{\leq K})} \\ &= \frac{\mathbb{P}(x(T) = R_i, \phi^{\leq K} = x^{\leq K})}{\sum_{j=1}^{N} \mathbb{P}(x(T) = R_j, \phi^{\leq K} = x^{\leq K})}. \end{aligned} \tag{8}$$

We note that the denominator in (8) equals 0 with respect to a T-dimensional Lebesgue measure. Namely, the set of all paths such that $\phi^{\leq K} = x^{\leq K}$ is of a lower dimension than the set of all paths of length T. Nonetheless, we can define the probability mass of all paths to R_j of length T such that $\phi^{\leq K} = x^{\leq K}$ from (7), and integrate it with respect to a $T-K$-dimensional Lebesgue measure. We obtain

$$P_i = \frac{\prod_{k=1}^{n} v^{T-K}(R_{ik} - x_k(K))/v^T(R_{ik})}{\sum_{j=1}^{N} \prod_{k=1}^{n} v^{T-K}(R_{jk} - x_k(K))/v^T(R_{jk})}. \tag{9}$$

We note the analogy between (6) and (9): instead of considering the cardinality of a set of paths as we did in Section III, we are now considering the volume of that set.

We now only need to compute $v^t(y)$ for $y \in \mathbb{R}$. An expression for $v^t(y)$ was recently provided [8]:

$$v^t(y) = \sum_{k=0}^{t} (-1)^k \binom{t}{k} \operatorname{sign}(t-2k+y) \frac{(t-2k+y)^{t-1}\sqrt{t}}{(t-1)!}. \tag{10}$$

Problem 3 is thus solved by applying (2), (9), and (10). Having solved this problem both in the case of arbitrary finite sets of allowed actions and for the continuous set $\mathcal{U}(x) = [-1, 1]^n$, we proceed to a brief discussion of optimal environment design.

V. Optimal Environment Design

We turn our attention to Problem 4. In this section we provide a globally optimal solution and a computationally feasible approximation method in the case of a finite set of allowed actions.

In Theorem 7 and the discussion in Section IV we provided an expression for target unpredictability given an agent's path until time K. By those results, entropy $H(P|x^{\leq K}, \mathcal{R})$ does not depend on all states $x(t)$ for $t \in [K]$, but only on $x(K)$. Hence, the solution to the problem of optimal environment design is the same as the solution to

$$\operatorname*{argmin}_{\mathcal{R}} \max_{y \in \overline{R}^{T-K}} H_K(P|y, \mathcal{R}) \tag{11}$$

where $\overline{R}^{T-K}$ denotes the set of all states $y \in \mathbb{R}^n$ which are both reachable from 0 in time K and from which at least one target in $\mathcal{R}$ is reachable in time $T-K$, and $H_K(P|y, \mathcal{R})$ denotes $H(P|x^{\leq K}, \mathcal{R})$ for any $x^{\leq K}$ such that $x(K) = y$.

Consider the scenario where all $\mathcal{U}(x)$ are finite. The sets from which the decision variables in (11) are chosen are then finite. Namely, given that there are finitely many paths from 0 to $\mathcal{R}$ of length T, set $\overline{R}^{T-K}$ is finite for every $\mathcal{R}$. On the other hand, we can consider only those $\mathcal{R}$ that consist of states that are reachable from 0 in time T. The set $\hat{R}^t$ of reachable states at time t satisfies $\hat{R}^t = \cup_{y \in \mathcal{R}^{t-1}} \{y + u \mid u \in \mathcal{U}(y)\}$ for all $t \in \mathbb{N}$, with $\hat{R}^0 = \{0\}$. Inductively, all sets $\hat{R}^t$, including $\hat{R}^T$, are finite.

Optimization problem (11) can now be solved exactly by computing all entropies H_K. If M is the number of states that can be reached from 0 in time no more than T, the number of entropies H_K that need to be calculated is bounded from above by M^{N+1}. In addition, computing each entropy from (5) and (6) requires up to $2NTM^2$ additions to generate values $p^t(y; y_E)$; some values can also be memorized instead of recomputed. While the computation time is thus exponential in the number of targets N, we note that the number of targets in deception and goal recognition scenarios [6], [13], [17], [22] is usually small: often, $N < 5$.

For larger N we propose the following method of finding a local optimum. After determining $M_{\mathcal{R}} = \max_y H_K(P|y, \mathcal{R})$ for a particular environment $\mathcal{R}$, compare

$M_{\mathcal{R}}$ to $H_K(P|y,\mathcal{R}')$ for all environments $\mathcal{R}'$ obtained from $\mathcal{R}$ by moving one target to a neighboring state. If $\max_y H_K(P|y,\mathcal{R}') < M_{\mathcal{R}}$, move $\mathcal{R}$ accordingly and continue. Algorithm 8 formally presents the method.

Algorithm 8:

Select an initial guess for set of targets $\mathcal{R}$.
Compute $M_{\mathcal{R}} = \max_{y \in \overline{\mathcal{R}}^{T-K}} H_K(P|y,\mathcal{R})$.
repeat until break
  for $j = 1, \ldots, N$ and all $u \in \mathcal{U}(R_j)$
    Define $\mathcal{R}^{(ju)}$ by $R_i^{(ju)} = R_i$ for $j \neq i$, $R_j^{(ju)} = R_j + u$.
    Compute $M_{\mathcal{R}}^{(ju)} = \max_{y \in \overline{\mathcal{R}}^{T-K}} H_K(P|y,\mathcal{R}^{(ju)})$.
  end for
  if any $M_{\mathcal{R}}^{(j^*u^*)} < M_{\mathcal{R}}$, **then**
    Define $\mathcal{R} = \text{argmin}_{j,u} \mathcal{R}^{(ju)}$.
  else
    break
  end if

Assuming that all the entropy values are stored, Algorithm 8 is, in the worst case, at least as fast as an exhaustive search. However, each run of the repeat loop requires computation of the number of entropies linear in N; since each entropy is calculated in $O(2NTM^2)$ operations, each run of the repeat loop is quadratic in N. Additionally, the algorithm obviously yields a locally optimal environment, in the sense that no single move of a target will produce better results. In practice, as we will show in the subsequent section, Algorithm 8 produces nearly optimal environments in significantly shorter time than the global algorithm.

The above methods provide both local and global optima for the problem of optimal environment design in the case of finite sets of actions. The same problem is significantly more challenging in the case of continuous set of actions $\mathcal{U} = [-1,1]^n$ considered in Section IV. Namely, (11) is a minimax problem over a continuous domain in $\mathbb{R}^{n\times(N+1)}$, with an additional difficulty that, while the objective function can be computed from (2), (9), and (10), it is nonlinear and, if differentiable, its derivative would be difficult to obtain analytically. In light of these features, a discretization to the case of a finite state space as discussed above might be the most meaningful approach.

We now proceed to validate our solution to Problem 3 and the subsequent solution of Problem 4 in a finite state space.

VI. Numerical Evaluation

The case study in this section builds on the model proposed in [17]. In it, an agent moves one tile north, east, south, west, or stays in place at each time step, operating on an 8×8 rectangular grid with three potential targets. The observer predicts the agent's true target at every time step.

A feature of the scenario from [17] that deviates from the assumptions of our paper is that the agent has incentive to be explicitly deceptive as opposed to just unpredictable: instead of a simple reachability objective, the agent in [17] receives a reward if it visits its intended target without the observer realizing that this particular target is the agent's intended one. The agent incurs a cost, i.e., receives a negative reward, if it visits the intended target while the observer does at that time believe that it is the agent's intended target.

We purposefully choose to test our results on the model of [17], having in mind that it differs from the model of our paper. The motivation that underlies our technical approach is the desire to protect against deception without knowing the deceptive agent's strategy or the observer's belief evolution. We will thus show that for the particular model of the agent's and observer's behaviors used in [17], even though those differ from the ones described in Assumption 1 and Assumption 2, the observer performs significantly better in the environment deemed to be optimal by our paper than in the environment from [17].

We omit the details of the parameters, optimal deceptive policies, and observer's target prediction from [17]. For broad understanding, we remark that the observer does not calculate a target prediction vector in the way proposed in our paper. Instead, it changes its prediction stochastically at every time depending on the agent's last step (i.e., if the agent moves closer to a potential target R_i, the observer will, with some probability, switch to believing that the true target R equals R_i). The agent's optimal deceptive policy is simply a policy that attains the maximal expected reward, assuming that the agent knows the observer's prediction at every time step.

The environment, i.e., the target placement, considered in [17] is illustrated in blue in Fig. 3. In this environment, when the agent uses its optimal deceptive policy, the observer correctly predicts the agent's target only around 21% of the time, based on 100 repetitions of 200 time steps each. In other words, the agent is truly deceptive: even though there are only three potential targets, an observer makes an incorrect prediction almost 80% of the time.

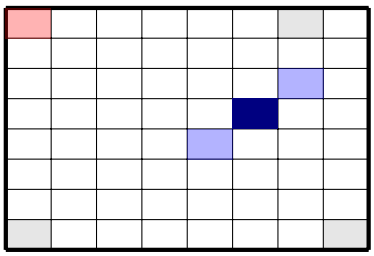

Fig. 3. An illustration of the environment considered in [17]. Potential targets $\mathcal{R}$ considered in [17] are denoted in blue; the agent's true target is darker, while the other two targets are lighter. The agent's initial state is denoted in red. The computed optimal environment is denoted in gray.

We now compute the optimal environment using the approach of Section V. We assume that the observer does not correctly know the agent's length of path, computing the optimal environment for $T = 20$ instead of 200. We also consider $K = 10$, effectively assuming that the observer solely sees the agent's position at time $t = 10$ instead of at every time step. An optimal environment is given in gray in Fig. 3; while targets are more distant than the ones used

in [17], we note that they are not chosen to be the states most distant from each other. Using the new environment, the observer correctly predicts the agent's true target around 53% of time: even though the agent is attempting to be deceptive, the observer mostly predicts the agent's target correctly.

While our results confirm the strong role of the environment in detection of deception, we do not guarantee that the optimal environment in the sense of Section V is optimal *for the particular agent's policy and observer's prediction method* used in [17]. In fact, a locally optimal environment computed using Algorithm 8 may produce marginally better results than the one obtained using exhaustive search; depending on the randomly chosen initial environment, the resulting environment of Algorithm 8 ensured that the observer correctly predicts the agent's true target $50 - 54\%$ of time. To illustrate the computational value of Algorithm 8, we note that computing the globally optimal environment took approximately 650 times longer than computing a locally optimal environment by Algorithm 8. The length of the latter computation was on the order of milliseconds.

Finally, to illustrate the results of Section IV, we compare target entropies $H(P|x^{\leq 10}, \mathcal{R})$ for the two environments in Fig. 3, this time assuming that the agent's actions come from a continuous set as described in Section IV. We again take $T = 20$, $K = 10$, and randomly with a uniform distribution generate 100 possible states y inside the rectangle $[0, 7] \times [0, 7]$. The maximal entropy $H_K(P|y, \mathcal{R})$ for the blue environment, used in [17], equals 1.1. The maximal entropy for the gray environment equals 0.87, confirming that the newly proposed environment is better than the one used in [17] in the case of continuous sets of actions.

VII. Conclusion and Future Work

This paper presents an initial discussion of the role of environment design in counterdeception. By taking a classical scenario of an agent seeking to reach one of multiple potential targets, we proposed a metric of unpredictability of the agent's target from its partial path *given the placement of potential targets*. This metric paves the way towards counterdeceptive environment design. In a representative example, by comparing a deceiving agent's success in a previously considered naively chosen environment and in a newly computed counterdeceptive environment, we showed that the latter indeed significantly reduces deceiver's abilities.

Obviously, much remains in the line of investigation that the paper seeks to initiate. Namely, the system dynamics that we assume in the paper are simple, and the assumption that the agent chooses its path entirely at random often does not hold in practice. Additionally, the observer might have some, if likely vague, information about the agent's strategy. While the idea of constructing the probability distribution on the set of agent's paths remains meaningful for more involved dynamics, computing these probabilities is correspondingly more difficult. Previous work on deception also naturally considers agents moving in continuous time: while the broad approach of defining a target prediction vector (2) again makes sense in such a setting, determining the probability mass of the underlying sets of paths presents a problem in measure theory on infinite-dimensional spaces.

Finally, in this paper we merely scratched the surface of optimal environment design; while we solve the problem for agents that move on a discrete state space, optimization in the case of a continuous state space, discussed in Section IV, calls for more advanced tools.

VIII. Acknowledgments

The author thanks Mustafa Karabag for a helpful discussion on Section III.

References

[1] A. Ayub, "An adaptive Markov process for deceptive robotics," Master's thesis, Pennsylvania State University, 2017.

[2] G. Chen, D. Shen, C. Kwan, J. B. Cruz, M. Kruger, and E. Blasch, "Game theoretic approach to threat prediction and situation awareness," *Journal of Advances in Information Fusion*, vol. 2, no. 1, pp. 35–48, 2007.

[3] S. H. Choi, "Enumeration of NSEW-paths in restricted planes," *Journal of the Korean Mathematical Society*, vol. 33, no. 2, pp. 413–421, 1996.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 2006.

[5] J. De Loera, J. Rambau, and F. Santos, *Triangulations: Structures for Algorithms and Applications*. Springer, 2010.

[6] A. Dragan, R. Holladay, and S. Srinivasa, "Deceptive robot motion: synthesis, analysis and experiments," *Autonomous Robots*, vol. 39, pp. 331–345, 2015.

[7] E. Erkut and S. Neuman, "Analytical models for locating undesirable facilities," *European Journal of Operational Research*, vol. 40, no. 3, pp. 275–291, 1989.

[8] R. Frank and H. Riede, "Hyperplane sections of the n-dimensional cube," *American Mathematical Monthly*, vol. 119, no. 10, pp. 868–872, 2012.

[9] P. R. Halmos, *Measure Theory*. Springer, 1978.

[10] S. Hesse Owen and M. S. Daskin, "Strategic facility location: A review," *European Journal of Operational Research*, vol. 111, no. 3, pp. 423–447, 1998.

[11] J.-Y. Jian, T. Matsuka, and J. V. Nickerson, "Recognizing deception in trajectories," in *28th Annual Conference of the Cognitive Science Society*, 2006, pp. 1563–1568.

[12] M. O. Karabag, M. Ornik, and U. Topcu, "Optimal deceptive and reference policies for supervisory control," in *58th IEEE Conference on Decision and Control*, 2019, pp. 1323–1330.

[13] P. Masters, "Goal recognition and deception in path-planning," Ph.D. dissertation, RMIT University, 2019.

[14] P. Masters and S. Sardina, "Deceptive path-planning," in *26th International Joint Conference on Artificial Intelligence*, 2017, pp. 4368–4375.

[15] ——, "Cost-based goal recognition for the path-planning domain," in *27th International Joint Conference on Artificial Intelligence*, 2018, pp. 5329–5333.

[16] ——, "Cost-based goal recognition in navigational domains," *Journal of Artificial Intelligence Research*, vol. 64, pp. 197–242, 2019.

[17] M. Ornik and U. Topcu, "Deception in optimal control," in *56th Annual Allerton Conference on Communication, Control, and Computing*, 2018, pp. 821–828.

[18] Y. Savas, M. Ornik, M. Cubuktepe, M. O. Karabag, and U. Topcu, "Entropy maximization for Markov decision processes under temporal logic constraints," *IEEE Transactions on Automatic Control*, 2019.

[19] J. Shim and R. C. Arkin, "Biologically-inspired deceptive behavior for a robot," in *12th International Conference on Simulation of Adaptive Behavior*, 2012, pp. 401–411.

[20] United States Department of Defense, "Military handbook design guidelines for physical security of facilities," Tech. Rep. ADA320793, 1993.

[21] ——, "Joint security operations in theater," Tech. Rep. JP 3-10, 2019.

[22] K. Xu, Y. Zeng, L. Qin, and Q. Yin, "Single real goal, magnitude-based deceptive path-planning," *Entropy*, vol. 22, no. 1, 2020.