Ethically Compliant Autonomous Systems under Partial Observability

Qingyuan Lu[1], Justin Svegliato[2], Samer B. Nashed[3], Shlomo Zilberstein[3], and Stuart Russell[2]

***Abstract*— Ethically compliant autonomous systems (ECAS) are the prevailing approach to building robotic systems that perform sequential decision making subject to ethical theories in fully observable environments. However, in real-world robotics settings, these systems often operate under partial observability because of sensor limitations, environmental conditions, or limited inference due to bounded computational resources. Therefore, this paper proposes a partially observable ECAS (PO-ECAS), bringing this work one step closer to being a practical and useful tool for roboticists. First, we formally introduce the PO-ECAS framework and a MILP-based solution method for approximating an optimal ethically compliant policy. Next, we extend an existing ethical framework for prima facie duties to belief space and offer an ethical framework for virtue ethics inspired by Aristotle's Doctrine of the Mean. Finally, we demonstrate that our approach is effective in a simulated campus patrol robot domain.**

I. Introduction

As autonomous systems take an increased role in society and individuals' well-being, it is important that they operate ethically. A promising approach to this challenge seeks to implement ethical theories within autonomous systems in order to leverage the extensive work of moral philosophers, which mirrors the considerations of users and lawmakers. Understandably, this is a hard task due to the tension between the ambiguity inherent to ethical theories and the precision demanded by autonomous systems. Naturally, one might try to implement an ethical theory within an autonomous system by directly modifying its objective function. However, modifying this objective function often results in unpredictable behavior by introducing an incommensurable trade-off between task completion and ethical compliance [24].

Ethically compliant autonomous systems [24], [20] are the prevailing approach to implementing ethical theories within autonomous systems. To do this, *ECAS* represents a decision-making problem (i.e., a Markov decision process) as a mathematical program with an additional ethical constraint that encodes an ethical theory. Hence, solving this mathematical program results in a policy that is guaranteed to comply with the ethical theory. However, while ECAS has been shown to be effective, it is limited to fully observable environments and cannot be used in partially observable environments. This is often critical for real-world robotics settings because of sensor limitations, environmental conditions, or limited inference due to bounded computational resources.

[1]Massachusetts Institute of Technology, Cambridge, MA, USA. Email: {kqlu@mit.edu}. [2]University of California, Berkeley, CA, USA. Emails: {jsvegliato, russell}@berkeley.edu. [3]University of Massachusetts, Amherst, MA, USA. Emails: {snashed, shlomo}@cs.umass.edu. Partial support for this work was provided by NSF grants 1954782 and 2205153.

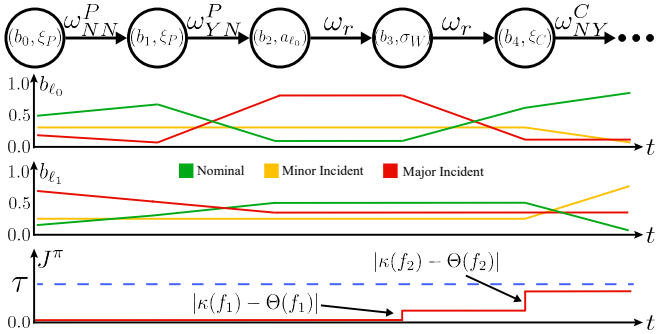

Fig. 1. A campus patrol robot monitors different security feeds and resolves incidents across locations ℓ_0 and ℓ_1. *Top:* The trajectory of the POMDP with belief-action pairs (b, a) connected by observations ω. Here, the action ξ_P monitors MAJOR incidents with a police scanner, ξ_C monitors MINOR incidents with a camera system, a_ℓ navigates to the location ℓ, and σ_W resolves an incident by warning campus security. Each observation ω is given by a security feed (P or C) and an incident status for the locations ℓ_0 and ℓ_1 (Y indicates an incident and N indicates no incident). ω_r indicates a destination has been reached. *Center:* The agent's belief b_ℓ for MINOR and MAJOR incidents at locations ℓ_0 and ℓ_1. *Bottom:* The cumulative penalty J^π of a policy π due to performing actions outside of the virtuous means $\Theta(f)$ given each virtue f (*prudence*/*courage*) for a strength function $\kappa(f)$ that reflects where along the extremes an action falls and a tolerance τ that allows for deviations from the virtuous mean.

Therefore, in this paper, we propose a novel approach to building *partially observable ECAS*. In particular, starting with the mathematical program for a decision-making problem in a partially observable environment (i.e., a partially observable Markov decision process), we formally represent *PO-ECAS* as a mixed integer linear program (MILP) with an additional ethical constraint that encodes an ethical theory. Importantly, in our experiments, we demonstrate the effectiveness of our approach in a simulated campus patrol robot domain in order to assess whether our approach can correctly, reliably, and predictably adjust the operation of an autonomous system in a partially observable environment.

The effectiveness of ECAS-based approaches heavily depends on the expressiveness of the given ethical theory. To improve this expressiveness, we introduce an ethical theory for virtue ethics inspired by Aristotle's Doctrine of the Mean—which requires an agent to act virtuously according to virtues that fall as mean values between vices of excess and deficiency—as illustrated in Figure 1. Intuitively, this ethical constraint enables an autonomous system to flexibly modulate its operation while maintaining virtuous character.

Our main contributions in this paper are: (1) a formal definition of partially observable ethically compliant autonomous systems and a MILP-based solution method, (2) an extension of an existing ethical framework for prima facie duties to belief space, (3) an ethical framework for virtue ethics based on Aristotle's Doctrine of the Mean, and (4) a demonstration of our approach in a simulated campus patrol robot domain.

II. Related Work

The application of ethical reasoning to automated systems at conception, regulation, design, and deployment is a broad and nuanced field of research. This paper focuses on ethical reasoning during stochastic planning, specifically for sequential decision-making models. Readers seeking a holistic treatment of this literature are directed to the ECAS framework [24], [20] that we build on in this paper along with related surveys [10], [19], [29], [28]. Most work that constrains decision making to follow an ethical theory uses logic systems [26], [7], [27], [3], including some that reason over a set of logics [8] or use Answer Set Programming [5].

Logic systems have several benefits, including their interpretability and their accessibility to theoretical tools and guarantees. However, these systems present a drawback: nuanced behavior can become difficult to specify as the capability of an agent grows, and deploying such systems in stochastic environments presents still unsolved challenges [1]. As a result, there is work that models ethical behavior using other mechanisms, such as game-theoretic concepts [13] and semantic orderings over logical statements as in Belief-Desire-Intention architectures [12], [11], case-supported principle-based behavior models [2], or multi-coherence theory [28]. Moreover, there are approaches that combine elements of rule-based systems with human oversight [9].

Still, these approaches pose two problems. First, they do not result in guarantees as they blend task completion and ethical compliance into a single objective function. Second, they do not consider partial observability. The ECAS framework [24], [20] addresses the first problem well: it allows developers to separate task completion from ethical compliance by constraining the space of policies of the decision-making model with an ethical theory. However, the framework and solution methods provided by ECAS require the decision-making model to be a Markov decision process, which assumes full observability.

Work on ethically-aware reasoning under partial observability has received far less attention. Many problems are partially observable due to their multi-agent or decentralized nature, so much of existing research focuses on this area, for example signaling normative intent in multi-agent systems [14], [4] or decentralized decision-making for autonomous vehicles [25]. Recently, there has been work on the fairness-privacy tradeoff that arises in decentralized systems when sharing information [22]. Most similar is work on the "resilience" of decentralized partially observable systems, modeled as hidden Markov models, where resilience is defined in terms of different metrics related to the expected cost of executing certain trajectories [21].

In summary, existing approaches fail to offer a general framework for decision-making, ethical constraints, and solution methods that can be applied to ethical reasoning in partially observable, stochastic environments. Moreover, in most approaches, the task objective cannot be cleanly separated from any ethical constraints, which is the primary advantage of the original ECAS framework.

III. Background

A *partially observable Markov decision process* (POMDP) is a formal decision-making model for reasoning in partially observable, stochastic environments [15]. A POMDP is a tuple $\langle S, A, T, R, \Omega, O \rangle$. S is a set of states of the world. A is a set of actions of the agent. $T : S \times A \times S \to [0, 1]$ is a transition function that maps each state $s \in S$ and action $a \in A$ to the probability of ending up in state $s' \in S$. $R : S \times A \to \mathbb{R}$ is a reward function that maps each state $s \in S$ and action $a \in A$ to the expected immediate reward. Ω is the set of observations of the agent. $O : S \times A \times \Omega \to [0, 1]$ is the observation function that maps each state $s \in S$ and action $a \in A$ to the probability of emitting observation $\omega \in \Omega$.

In a POMDP, the agent does not necessarily know the true state of the world at any given time. Instead, the agent makes noisy observations that reflect its state and action. To represent its uncertainty, the agent maintains a belief state $b \in B$, a probability distribution over all states, where B is the space of all belief states. Initially, the agent begins with an initial belief state $b_0 \in B$. After performing an action $a \in A$ and making an observation $\omega \in \Omega$, the agent updates its current belief state $b \in B$ to a new belief state $b' \in B$ using the belief state update equation $b'(s'|b, a, \omega) = \alpha O(a, s', \omega) \sum_{s \in S} T(s, a, s') b(s)$, where α is the normalization constant $\alpha = Pr(\omega|b, a)^{-1}$.

A policy π of a POMDP can be represented as a *finite-state controller* (FSC) of a fixed size. Formally, an FSC is a tuple $\pi = \langle Q, \lambda, \eta \rangle$. Q is a set of nodes that each represent a region of the belief space B. $\lambda : Q \to A$ is an action function that maps each node $q \in Q$ to an action $a \in A$. $\eta : Q \times \Omega \to Q$ is a transition function that maps each node $q \in Q$ and observation $\omega \in \Omega$ to a successor node $q' \in Q$. At each time step, the agent begins in a node $q \in Q$ associated with its current belief state $b \in B$, performs an action $a \in A$ given the action function λ, and ends up in a successor node $q' \in Q$ given the transition function η. An FSC π induces a value function $V^\pi : Q \times S \to \mathbb{R}$ that represents the expected cumulative reward of a node $q \in Q$ and a state $s \in S$. Naturally, an optimal FSC π^* maximizes this value function. Note that solution methods that yield FSCs of a fixed size may not be optimal because they restrict the space of policies [6].

IV. Operating under Partial Observability

In existing work, ECAS has operated under full observability [24], [20]. However, full observability is often an impractical assumption in real-world robotics settings. For instance, consider a robot that must patrol different locations on a college campus to resolve potential incidents. Naturally, it has uncertainty over the location and severity of potential incidents but can monitor different feeds that provide information, such as a camera system or a police scanner. More generally, in the real world, robots often only make *observations* that affect their *belief* about the world. As a result, managing this belief subject to sensor limitations, environmental conditions, and limited inference coupled with the goal of completing a task efficiently and reliably in

a stochastic, unstructured environment makes this class of decision-making problems ethically challenging.

A *partially observable ethically compliant autonomous system* (PO-ECAS) has a POMDP as the decision-making model for completing its *task* and an ethical context and a moral principle for following its *ethical framework* [24]. The POMDP $\mathcal{P}$ describes the information needed to complete the task, the ethical context $\mathcal{E}$ describes the information required to follow the ethical framework, and the moral principle $\rho : \Pi \to \mathbb{B}$ evaluates the morality of a policy for the decision-making model within the ethical context. We formally describe a PO-ECAS in the following way:

Definition 1. *A **PO-ECAS** $\langle \mathcal{P}, \mathcal{E}, \rho \rangle$ optimizes completing a task by using a POMDP $\mathcal{P}$ while following an ethical framework by adhering to a moral principle $\rho : \Pi \to \mathbb{B}$ within an ethical context $\mathcal{E}$. $\mathbb{B}$ is the Boolean set $\{0, 1\}$.*

The objective of a PO-ECAS is to find an optimal policy that maximizes the expected cumulative reward of the POMDP subject to following the ethical framework:

Definition 2. *The objective of a PO-ECAS is to find an **optimal moral policy**, $\pi^*_\rho \in \Pi$, by solving for a policy $\pi \in \Pi$ within the space of policies Π that maximizes a value function V^π subject to a moral principle ρ:*

$$\underset{\pi \in \Pi}{\textit{maximize}} \quad V^\pi \quad \textit{subject to} \quad \rho(\pi)$$

A PO-ECAS can follow an ethical framework that impacts completing its task. This impact can be measured as the maximum absolute difference across all states between the value function of the optimal moral policy and the value function of the optimal amoral policy:

Definition 3. *Given the optimal moral policy $\pi^*_\rho \in \Pi$ and the optimal amoral policy $\pi^* \in \Pi$, the **price of morality**, ψ, can be represented by the expression $\psi = \|V^{\pi^*_\rho} - V^{\pi^*}\|_\infty$.*

It is possible to solve a PO-ECAS approximately by using mathematical programming. Table I offers a novel mixed integer linear program (MILP) for a POMDP that is based on existing work on DecPOMDPs [17]. Unlike an MDP, which can be expressed as a linear program that generates a deterministic or stochastic policy depending on whether or not there are constraints, a POMDP is expressed as a MILP that generates a deterministic policy. A deterministic policy of the MILP is represented as a deterministic FSC of a fixed size for the POMDP. Deterministic FSCs are a standard representation for policies in work on POMDPs [15].

In our MILP for POMDPs, the variables consist of (1) *occupancy measures* that represent the frequency that the FSC performs an action in a node and a state and (2) *probabilities* that represent the FSC's action/transition functions. The objective maximizes the discounted cumulative reward given the FSC's occupancy measures $x(q, s, a)$. Flow Consistency Constraint 1 ensures that the FSC's occupancy measures are consistent given the observation and transition functions of the POMDP and the initial node and state distribution η_0. Probability Constraints 2-5 ensure that the FSC's occupancy measures are valid marginalizations. Probability Constraints 6-7 ensure that the FSC's action/transition functions are valid. Probability Constraints 8-9 ensure that the FSC's action/transition functions are probability distributions. Integrality Constraints 10-11 ensure that the FSC's action/transition functions are deterministic. Note that $x(a|q)$ and $x(q'|q, o)$ represent the FSC's action/transition function and the remaining variables are occupancy measures under different marginalizations. For a detailed explanation of this form of MILPs, see existing work on DecPOMDPs [17].

TABLE I

A MIXED INTEGER LINEAR PROGRAM REPRESENTATION OF A POMDP.

Variables:

$$x(q, s, a), x(q, s, a, q'_\omega), x(q, a), x(q), x(q, q'_\omega), x(a|q), x(q'|q, \omega) \quad \forall q, s, a, \omega, q'$$

Maximize:

$$\sum_{s \in S} \sum_{a \in A} R(s, a) \sum_{q \in Q} x(q, s, a)$$

Flow Consistency Constraint:

$$\sum_{a \in A} x(q', s', a) = \eta_0(q', s') + \gamma \sum_{s \in S} \sum_{a \in A} \sum_{\omega \in \Omega} O(a, s', \omega) T(s, a, s') \sum_{q \in Q} x(q, s, a, \tilde{q}_\omega = q') \ \forall q', s' \quad (1)$$

Probability Constraints:

$$x(q, s, a) = \sum_{q' \in Q} x(q, s, a, q'_\omega) \ \forall q, s, a, \omega \quad (2)$$

$$x(q, a) = \sum_{s \in S} x(q, s, a) \ \forall q, a \quad (3)$$

$$x(q) = \sum_{a \in A} x(q, a) \ \forall q \quad (4)$$

$$x(q, q'_\omega) = \sum_{s \in S} \sum_{a \in A} x(q, s, a, q'_\omega) \ \forall q, \omega, q' \quad (5)$$

$$x(q) - x(q, a) \leq \frac{1 - x(a|q)}{1 - \gamma} \ \forall q, a \quad (6)$$

$$x(q) - x(q, q'_\omega) \leq \frac{1 - x(q'|q, \omega)}{1 - \gamma} \ \forall q, \omega, q' \quad (7)$$

$$\sum_{a \in A} x(a|\omega) = 1 \ \forall q \quad (8)$$

$$\sum_{q' \in Q} x(q'|q, \omega) = 1 \ \forall q, \omega \quad (9)$$

Integrality Constraints:

$$x(a|q) \in \{0, 1\} \quad (10)$$

$$x(q'|q, \omega) \in \{0, 1\} \quad (11)$$

In general, while MILPs are NP-hard, reasonably sized MILPs can be solved using modern processors and methods. Our experiments execute the default simplex method in the IBM CPLEX Optimization Suite using a MIP gap of 0.1, an integrality tolerance of 0.005, and a discount factor γ of 0.99 on a MacBook Pro with an M1 CPU and 16 GB RAM.

V. ETHICAL FRAMEWORKS IN BELIEF SPACE

In this section, we extend an existing ethical framework for prima facie duties to belief space and offer an ethical framework for virtue ethics based on Aristotle's Doctrine of the Mean. Table II provides the ethical constraints that are added to the MILP for each ethical framework. *Conjunctions* is the number of logical conjunctions. *Operations* is an upper bound on the number of mathematical operations. *Computations* is an upper bound on the number of computations.

a) Prima Facie Duties: This is a pluralistic, nonabsolutist ethical theory that states that the morality of an action is based on whether that action fulfills fundamental moral duties that can contradict each other [23], [18]. We approximate prima facie duties (PFD) by formalizing an ethical framework that requires a policy to select actions that do not neglect duties of different penalties within a tolerance.

TABLE II

THE ETHICAL CONSTRAINTS REPRESENTING THE MORAL PRINCIPLES OF EACH ETHICAL FRAMEWORK WHERE $|T| = |S||A||S|$.

Moral Constraint	Observability	Conjunctions	Operations	Computations
$c^{\text{Fo}}_{\rho_\Delta}(\mu) = \sum_{s\in S, a\in A} \mu^s_a \sum_{\delta\in\Delta_{s,a}} \phi(\delta, s, a) \leq \tau$	Full	1	$2\|T\|\|\Delta\| + 1$	$2\|T\|\|\Delta\| + 1$
$c^{\text{Po}}_{\rho_\Delta}(\mu) = \sum_{s\in S, a\in A, q\in Q} \mu^{s,q}_a \sum_{\delta\in\Delta_{s,a}} \phi(\delta, s, a) \leq \tau$	Partial	1	$2\|Q\|\|T\|\|\Delta\| + 1$	$2\|Q\|\|T\|\|\Delta\| + 1$
$c^{\text{Fo}}_{\rho_\mathcal{F}}(\mu) = \bigwedge_{f\in\mathcal{F}} \sum_{s\in S, a\in A} \mu^s_a \lvert\kappa(f, s, a) - \Theta(f)\rvert \leq \tau$	Full	$\lvert\mathcal{F}\rvert$	$2\|T\| + 1$	$\lvert\mathcal{F}\rvert(2\|T\| + 1)$
$c^{\text{Po}}_{\rho_\mathcal{F}}(\mu) = \bigwedge_{f\in\mathcal{F}} \sum_{s\in S, a\in A, q\in Q} \mu^{s,q}_a \lvert\kappa(f, s, a) - \Theta(f)\rvert \leq \tau$	Partial	$\lvert\mathcal{F}\rvert$	$2\|Q\|\|T\| + 1$	$\lvert\mathcal{F}\rvert(2\|Q\|\|T\| + 1)$

Definition 4. *A **PFD ethical context** $\mathcal{E}_\Delta$ is represented by a tuple $\mathcal{E}_\Delta = \langle \Delta, \phi, \tau \rangle$:*

- *Δ is a set of **duties**.*
- *$\phi : \Delta \times S \times A \to \mathbb{R}^+$ is a **penalty function** that represents the expected immediate penalty $\phi(\delta, s, a)$ for neglecting a duty $\delta \in \Delta$ when performing an action $a \in A$ in a state $s \in S$.*
- *$\tau \in \mathbb{R}^+$ is a **tolerance**.*

Definition 5. *A **PFD moral principle** ρ_Δ is expressed as:*

$$\rho_\Delta(\pi) = \sum_{s\in S} b_0(s) J^\pi(q_0, s) \leq \tau.$$

*The **expected cumulative penalty**, $J^\pi : Q \times S \to \mathbb{R}$, is:*

$$J^\pi(q, s) = \sum_{q'\in Q} \sum_{s'\in S} \bar{T}^{ss'}_{qq'} \big[\sum_{\delta\in\Delta_{s'}} \phi(\delta, s', \pi(q)) + J^\pi(q', s') \big],$$

where we let $\bar{T}^{ss'}_{qq'} = \bar{T}(\langle q, s\rangle, \pi(q), \langle q', s'\rangle)$.

This POMDP formulation is derived from the MDP formulation of prima facie duties in existing work [24]. In the moral principle, the initial belief distribution $b_0(s)$ of the POMDP formulation replaces the initial state distribution $d_0(s)$ of the MDP formulation. Multiplied by the initial belief distribution $b_0(s)$, the expected cumulative penalty $J^\pi(q_0, s)$ is parameterized by both an initial controller node q_0, representing a region of belief space, and a state s instead of only a state s. Moreover, in the expected cumulative penalty, a sum over both controller nodes q and states s replaces a sum over only states s, and the transition function represents transitions between both controller nodes q and states s instead of only states s. Note that if a belief MDP formulation—equivalent to the POMDP formulation—were used to represent prima facie duties, each state s would simply be replaced with a belief state b. The POMDP formulation is as expressive as the belief MDP formulation but more naturally supports approximate policies like FSCs that are commonly used in MILP-based solution methods.

b) Aristotelian Virtue Ethics: This is a monistic, absolutist ethical theory that states that the morality of an action depends on whether it embodies a set of virtues. In particular, this formulation of virtue ethics is based on Aristotle's Doctrine of the Mean and identifies every virtue as a condition that is *intermediate* or a "golden mean" falling between vices of excess and deficiency [16]. We approximate Aristotelian virtue ethics (AVE) by formalizing an ethical framework that requires a policy to select actions that yield a mean value between different vices of excess and deficiency within a tolerance.

Definition 6. *An **AVE ethical context** $\mathcal{E}_\mathcal{F}$ is represented by a tuple $\mathcal{E}_\mathcal{F} = \langle \mathcal{F}, \Theta, \kappa, \tau \rangle$:*

- *$\mathcal{F}$ is a set of **traits**.*
- *$\Theta : \mathcal{F} \to \mathbb{R}$ is a **virtue function** that represents the virtuous mean $\Theta(f)$ of a trait $f \in \mathcal{F}$.*
- *$\kappa : \mathcal{F} \times S \times A \to \mathbb{R}$ is a **strength function** that represents the strength $\kappa(f, s, a)$ of a trait $f \in \mathcal{F}$ when performing an action $a \in A$ in a state $s \in S$.*
- *τ is a **tolerance** for each trait $f \in \mathcal{F}$.*

Definition 7. *An **AVE moral principle** $\rho_\mathcal{F}$ is expressed as:*

$$\rho_\mathcal{F}(\pi) = \bigwedge_{f\in\mathcal{F}} \big[\sum_{s\in S} b_0(s) J^\pi_f(q_0, s) \leq \tau \big].$$

*The **expected cumulative penalty**, $J^\pi_f : Q \times S \to \mathbb{R}$, is:*

$$J^\pi_f(q, s) = \sum_{q'\in Q} \sum_{s'\in S} \bar{T}^{ss'}_{qq'} \big[\lvert\kappa(f, s', \pi(q)) - \Theta(f)\rvert + J^\pi_f(q', s') \big],$$

where we let $\bar{T}^{ss'}_{qq'} = \bar{T}(\langle q, s\rangle, \pi(q), \langle q', s'\rangle)$.

Although an MDP formulation for virtue ethics has been proposed in existing work [24], it differs considerably from the POMDP formulation introduced in this paper. Specifically, the moral principle here checks—for every trait f—that an agent starting with a belief b_0 associated with a controller node q_0 and following a policy π deviates no more than the tolerance τ in expectation from expressing that trait f at the virtuous mean $\Theta(f)$. This deviation at a state s' for a trait f is given by the norm $|\kappa(f, s', a) - \Theta(f)|$. Note that this ethical framework simplifies multiple concepts within virtue ethics and primarily serves as an example that demonstrates our approach but still retains rich descriptive and expressive power even with only a handful of ethical constraints.

To provide an example of this moral principle, consider the virtue *courage* that is often described as the golden mean between *cowardice* and *recklessness*. If pure cowardice and pure recklessness are represented by real numbers ζ_L and ζ_U respectively, the strength function $\kappa(f_{courage}, s, a)$ represents where within the extremes $[\zeta_L, \zeta_U]$ performing an action a in a state s falls. Accordingly, the virtue function $\Theta(f_{courage}) \in [\zeta_L, \zeta_U]$ represents the optimal tradeoff between the two extremes—the golden mean. Hence, any action a in which $\kappa(f_{courage}, s, a) \approx \Theta(f_{courage})$ would be courageous.

A PO-ECAS can define the moral principle in terms of a controller node q in an FSC, which covers a region of belief space. Naturally, this is an intuitive representation of a policy of a POMDP. However, unlike external factors such as the traits, virtue function, strength function, and

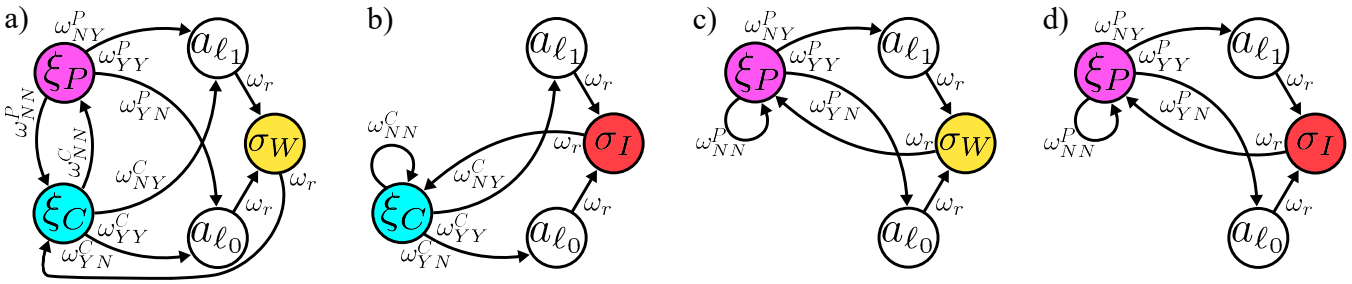

Fig. 2. The policies for the campus patrol robot domain generated for each version of AVE. These policies are illustrated as FSCs, where a *circle* denotes a controller node labeled with the action taken at that controller node and an *arrow* denotes a deterministic transition from one controller node to another controller node upon receiving the observation ω. Note that, in some policies, the observations ω^P_{YY} or ω^C_{YY}, indicating two simultaneous incidents of the same severity at different locations, can deterministically transition to an arbitrary choice of controller nodes.

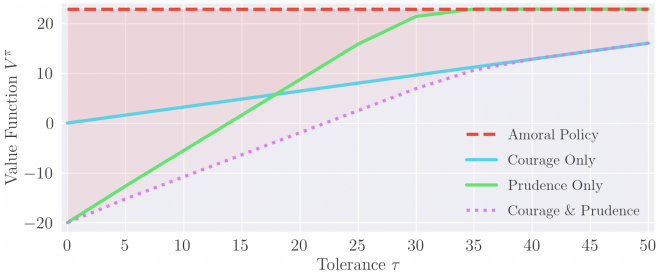

Fig. 3. The value functions for each version of AVE as a function of tolerance. Here, the *red dashed line* denotes the value function of the amoral policy that serves as an upper bound on the value function for each ethically compliant policy. The difference (shaded *red*) between the amoral policy and each ethically compliant policy denotes the price of morality.

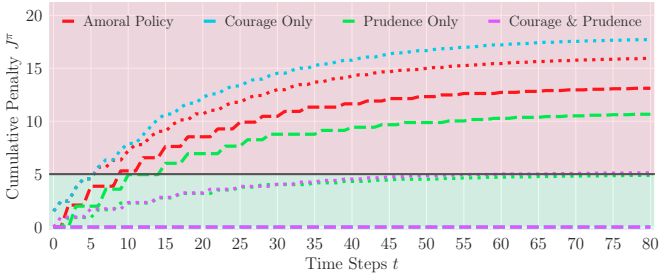

Fig. 4. The cumulative penalty functions for each version of AVE as a function of time steps. Here, the *dashed lines* denote the cumulative penalty for the *courage* trait and the *dotted lines* denote the cumulative penalty for the *prudence* trait. The *solid line* denotes a tolerance τ of 5 and separates the moral region (shaded *green*) and the immoral region (shaded *red*).

tolerance that are used in the moral principle, the agent can have a level of control over its belief depending on when and which information gathering actions it performs. Hence, under certain conditions, it may be possible for the agent to deliberately avoid collapsing its belief or otherwise alter its belief to avoid generating large penalties. We recognize that this issue raises ethical questions concerning whether or not ethical frameworks should embed a moral imperative to be maximally informed and leave this issue for our future work.

VI. Campus Patrol Robot Domain

We now turn to an application of PO-ECAS to a campus patrol robot. Here, the robot completes a campus patrol task that involves patrolling different locations on a college campus. During its patrol, it must resolve potential incidents by either intervening or warning campus security. It can make observations of the severity of a potential incident at each location by monitoring different security feeds at the base station. This means that the severity of each incident cannot be observed directly and evolves stochastically. Most importantly, the robot must follow the ethical framework for Aristotelian virtue ethics by operating courageously and prudently as discussed below. We describe both task completion and ethical compliance of the campus patrol robot below.

Task Completion

The robot must complete a campus patrol task that involves patrolling different locations L on a college campus. It begins at a base station L_B. Here, it can either navigate to a location $\ell \in L$ or monitor the severity $(\textsc{None}, \textsc{Minor}, \textsc{Major})$ of a potential incident at each location $\ell \in L$ using a camera system that noisily detects MINOR incidents or a police scanner that noisily detects MAJOR incidents. At each location $\ell \in L$, it can perform no operation ($\varnothing$) or resolve an incident by intervening (low effectiveness) or warning campus security (high effectiveness), which then returns it to the base station L_B. Overall, the objective of the robot is to resolve potential incidents across the college campus.

Formally, we represent the decision-making model of the campus patrol task by a POMDP $\mathcal{P} = \langle S, A, T, R, \Omega, O \rangle$. The set of states $S = \{\textsc{None}, \textsc{Minor}, \textsc{Major}\}^{|L|} \cup L \cup L_B$ represents all combinations of incident severities $\{\textsc{None}, \textsc{Minor}, \textsc{Major}\}^{|L|}$ for each location $\ell \in L$, and the current location $\ell \in L$ or base state L_B. The set of actions $A = A_L \cup A_\Sigma \cup A_\Xi \cup \varnothing$ has navigation actions $a_\ell \in A_L$ for navigating from the base station L_B to a location $\ell \in L$, resolution actions $A_\Sigma = \{\varnothing, \sigma_I, \sigma_W\}$ for performing no operation ($\varnothing$) and resolving an incident at a location $\ell \in L$ by intervening (σ_I) or warning campus security (σ_W), and monitoring actions $A_\Xi = \{\xi_C, \xi_P\}$ for using a camera system (ξ_C) to noisily detect MINOR incidents or a police scanner (ξ_P) to noisily detect MAJOR incidents. The set of observations $\Omega = \{\textsc{Yes}, \textsc{No}\}^{|L|} \cup \omega_r$ has observations $\{\textsc{Yes}, \textsc{No}\}^{|L|}$ for monitoring potential incidents for each location $\ell \in L$ and an arrival observation ω_r for reaching the base station L_B or a location $\ell \in L$.

Moreover, the reward, transition, and observation functions R, T, and O represent the dynamics of the campus patrol task. The transition function T reflects that navigation actions A_L move the robot from the base station L_B to a location $\ell \in L$, resolution actions A_Σ resolve incidents by changing the severity of an incident at a location $\ell \in L$ to NONE,

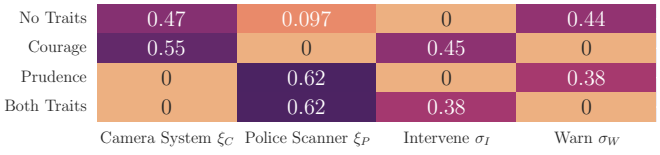

Fig. 5. A heat map of the action probabilities of the policies for the campus patrol robot domain generated for each version of AVE.

and monitoring actions A_Ξ that exogenously change the severity of an incident at a location $\ell \in L$. The reward function R reflects that navigation actions A_L yield nil reward, resolution actions A_Σ yield a small/large reward for resolving an incident of MINOR/MAJOR severity, and monitoring actions A_Ξ yield a small negative reward. The observation function O reflects that navigation actions A_L emit an arrival observation ω_r after navigating from the base station L_B to a location $\ell \in L$, resolution actions A_Σ emit an arrival observation ω_r after returning to the base station L_B from a location $\ell \in L$, and monitoring actions A_Ξ yield an observation $\omega \in \{\text{YES}, \text{NO}\}^{|L|}$ after monitoring the severity of potential incidents for each location $\ell \in L$.

Ethical Compliance

The robot must follow the ethical framework for Aristotelian virtue ethics with the *courage* and *prudence* traits f_1 and f_2. In any location state $\ell \in L$, the courage trait f_1 is *strengthened* (high κ_{f_1}) if the robot intervenes and *weakened* (low κ_{f_1}) if the robot performs no operation or warns the campus security to resolve an incident of any severity. In the base station L_B, the prudence trait f_2 is *strengthened* (high κ_{f_2}) if the robot checks the police scanner and *weakened* (low κ_{f_2}) if the robot checks the camera system to detect potential incidents. Thus, the robot must avoid the value of the courage and prudence traits f_1 and f_2 being higher/lower than its virtuous mean $\Theta(f_1)$ and $\Theta(f_2)$ for a tolerance τ.

Formally, given the AVE moral principle $\rho_\mathcal{F}$, we represent the AVE ethical context by a tuple $\mathcal{E}_\mathcal{F} = \langle \mathcal{F}, \Theta, \kappa, \tau \rangle$. The set of traits $\mathcal{F} = \{f_1, f_2\}$ represents the courage and prudence traits f_1 and f_2; the virtue function Θ represents the virtuous mean $\Theta(f)$ of the courage and prudence traits f_1 and f_2; the strength function κ represents the strength $\kappa(f, s, a)$ of courage and prudence traits f_1 and f_2 when performing an action a in a state s; and the tolerance τ allows for deviations from the courage and prudence traits f_1 and f_2.

VII. EXPERIMENTS

In our experiments, we evaluate the PO-ECAS approach in the campus patrol robot domain with Aristotelian virtue ethics for different combinations of traits: *no traits*, *courage*, *prudence*, and *both traits*. For each combination, we examine the final policy, value function, and cumulative penalty. To do this, after computing the final policy, we perform 100 simulations of the campus patrol robot domain with Aristotelian virtue ethics to obtain the value function and cumulative penalty. We focus on Aristotelian virtue ethics as existing work offers an analysis of prima facie duties [24].

a) Policy Analysis: Figure 2 shows the FSC policies generated for each version of Aristotelian virtue ethics (with the action probabilities across unobserved states in Figure 5). First, Figure 2a is the *amoral* policy (*no traits*) in which the agent performs both monitoring actions (ξ_C and ξ_P) but always warns campus security (σ_W) to resolve incidents due to its high effectiveness. Second, Figure 2b is the *courage* policy in which the agent only intervenes (σ_I) instead of warning campus security (σ_W) as $\kappa(f_{courage}, s, \sigma_I)$ is closer to $\Theta(f_{courage})$ than $\kappa(f_{courage}, s, \sigma_W)$. Third, Figure 2c is the *prudence* policy in which the agent only monitors the police scanner (ξ_P) instead of monitoring the camera system (ξ_C) as $\kappa(f_{prudence}, s, \xi_P)$ is closer to $\Theta(f_{prudence})$ than $\kappa(f_{courage}, s, \xi_C)$. Fourth, Figure 2d is the *courage and prudence* policy (*both traits*) that combines both the *courage* and *prudence* policy: the agent resolves incidents by intervening (σ_I) and only monitors the police scanner (ξ_P). Intuitively, despite being highly constrained, this policy is a natural combination of the policies in Figures 2b and 2c.

b) Value Function Analysis: Figure 3 shows the value function of the FSC policies generated for each version of Aristotelian virtue ethics over a range of tolerances. As expected, regardless of the combination of traits used in Aristotelian virtue ethics, the value function increases as the tolerance increases until reaching the upper bound (the value function of the amoral policy). This is because the tolerance allows the agent to smoothly change how much it follows the ethical theory. Moreover, the slope of these lines describes how the value function increases as the tolerance increases: when the value function slope is *steep* (*gentle*), actions that yield high reward on the task are *closer to* (*further from*) the virtuous mean for one or both traits. Hence, given just two values for the tolerance τ, we can predict the performance for each combination of traits. Lastly, as expected, *both traits* provide a lower bound on all value functions.

c) Cumulative Penalty Analysis: Figure 4 shows the cumulative penalty of the FSC policies generated for each version of Aristotelian virtue ethics over the time steps of the simulations. Here, the amoral policy is above the tolerance for *courage* and *prudence*. The *courage* policy is below the tolerance for *courage* and above the tolerance for *prudence*. The *prudence* policy is above the tolerance for *courage* and at the tolerance for *prudence*. The *courage and prudence* policy is below the tolerance for *courage* and at the tolerance for *prudence*. These observations are expected given the traits optimized for by each version of Aristotelian virtue ethics.

VIII. CONCLUSION

This paper proposes a novel framework for ethically compliant autonomous systems in partially observable environments and a MILP-based solution method. We then extend an existing ethical framework for prima facie duties to belief space and offer an ethical framework for virtue ethics based on Aristotle's Doctrine of the Mean. Finally, we show that our approach is effective in a simulated campus patrol robot domain. Future work will expand these ethical frameworks and develop more efficient approximate solvers.

REFERENCES

[1] D. Abel, J. MacGlashan, and M. L. Littman. Reinforcement learning as a framework for ethical decisions. In *AAAI Workshop on AI, Ethics, and Society*, 2016.

[2] M. Anderson and S. L. Anderson. Toward ensuring ethical behavior from autonomous systems: A case-supported principle-based paradigm. *Industrial Robot: An International Journal*, 42(4), 2015.

[3] K. Atkinson and T. Bench-Capon. Addressing moral problems through practical reasoning. In *International Workshop on Deontic Logic and Artificial Normative Systems*. Springer, 2006.

[4] C. Benn and A. Grastien. Reducing moral ambiguity in partially observed human–robot interactions. *Advanced Robotics*, 35(9), 2021.

[5] F. Berreby, G. Bourgne, and J.-G. Ganascia. Modelling moral reasoning and ethical responsibility with logic programming. In *Logic for Programming, Artificial Intelligence, and Reasoning*. Springer, 2015.

[6] D. Braziunas. POMDP solution methods. *University of Toronto*, 2003.

[7] S. Bringsjord, K. Arkoudas, and P. Bello. Toward a general logicist methodology for engineering ethically correct robots. *Intelligent Systems*, 22, 2006.

[8] S. Bringsjord, J. Taylor, B. Van Heuveln, K. Arkoudas, M. Clark, and R. Wojtowicz. Piagetian roboethics via category theory: Moving beyond mere formal operations to engineer robots whose decisions are guaranteed to be ethically correct. In *Machine Ethics*. Cambridge University Press, 2011.

[9] D. Brutzman, C. L. Blais, D. T. Davis, and R. B. McGhee. Ethical mission definition and execution for maritime robots under human supervision. *IEEE Journal of Oceanic Engineering*, 43(2), 2018.

[10] S. Cave, R. Nyrup, K. Vold, and A. Weller. Motivations and risks of machine ethics. *Proceedings of the IEEE*, 107(3), 2018.

[11] N. Cointe, G. Bonnet, and O. Boissier. Ethical judgment of agents' behaviors in multi-agent systems. In *15th International Conference on Autonomous Agents and Multiagent Systems*, 2016.

[12] N. Cointe, G. Bonnet, and O. Boissier. Multi-agent based ethical asset management. In *1st Workshop on Ethics in the Design of Intelligent Agents*, 2016.

[13] V. Conitzer, W. Sinnott-Armstrong, J. S. Borg, Y. Deng, and M. Kramer. Moral decision making frameworks for artificial intelligence. In *AAAI Workshop on AI, Ethics, and Society*, 2017.

[14] A. Grastien, C. Benn, and S. Thiébaux. Computing plans that signal normative compliance. In *2021 AAAI/ACM Conference on AI, Ethics, and Society*, 2021.

[15] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra. Planning and acting in partially observable stochastic domains. *Journal of AI Research*, 1998.

[16] R. Kraut. Aristotle's Ethics. In *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, 2022.

[17] A. Kumar, H. Mostafa, and S. Zilberstein. Dual formulations for optimizing Dec-POMDP controllers. In *International Conference on Planning and Scheduling*, 2016.

[18] M. Morreau. Prima facie and seeming duties. *Studia Logica*, 57(1), 1996.

[19] V. Nallur. Landscape of machine implemented ethics. *Science and Engineering Ethics*, 26, 2020.

[20] S. Nashed, J. Svegliato, and S. Zilberstein. Ethically compliant planning within moral communities. In *2021 AAAI/ACM Conference on AI, Ethics, and Society*, 2021.

[21] J. Panerati, N. Schwind, S. Zeltner, K. Inoue, and G. Beltrame. Assessing the resilience of stochastic dynamic systems under partial observability. *Plos One*, 13(8), 2018.

[22] A. Raymond, M. Malencia, G. Paulino-Passos, and A. Prorok. Agree to disagree: Subjective fairness in privacy-restricted decentralised conflict resolution. *Frontiers in Robotics and AI*, 9, 2022.

[23] W. D. Ross. *The right and the good*. Oxford University Press, 1930.

[24] J. Svegliato, S. B. Nashed, and S. Zilberstein. Ethically compliant sequential decision making. In *35th AAAI Conference on Artificial Intelligence*, 2021.

[25] B. Toghi, R. Valiente, D. Sadigh, R. Pedarsani, and Y. P. Fallah. Altruistic maneuver planning for cooperative autonomous vehicles using multi-agent advantage actor-critic. *arXiv preprint arXiv:2107.05664*, 2021.

[26] L. van der Torre. Contextual deontic logic: Normative agents, violations and independence. *Annals of Mathematics and Artificial Intelligence*, 37(5), 2003.

[27] M. Wooldridge and W. Van Der Hoek. On obligations and normative ability: An analysis of the social contract. *Journal of Applied Logic*, 3(4), 2005.

[28] L. Yilmaz, A. Franco-Watkins, and T. S. Kroecker. Computational models of ethical decision-making: A coherence-driven reflective equilibrium model. *Cognitive Systems Research*, 46, 2017.

[29] H. Yu, Z. Shen, C. Miao, C. Leung, V. R. Lesser, and Q. Yang. Building ethics into artificial intelligence. In *arXiv preprint arXiv:1812.02953*, 2018.