Extended Abstract: Mapping Target-Centered Virtue Ethics onto POMDPs for Value-Transparent Autonomous Systems

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—Autonomous agents inevitably enact social values in the ways they choose to distribute resources, communicate with other agents, and occupy space. However, in most autonomous systems, the way these values are implemented is opaque and non-configurable. We outline a value-configurable autonomous system through the novel mapping of target-centered virtue ethics to a partially observable MDP. We propose an application of the resulting system in a human-robot collaboration domain, where other agents' social states are estimated in order to both model social values and allow their configuration via a reward function.**

***Index Terms*—POMDP, virtue ethics, moral robots**

I. Introduction

Consider working alongside another person, where you both are assembling tool kits from a shared pool of tools. Your coworker cares only about finishing their task as fast as possible, and as such they consistently claim tools that you need for your task before you can access them. From an engineering perspective, this behavior is simply the result of optimizing for task efficiency. However, a human observer may interpret this behavior as inconsiderate or greedy, an emergent value-based attribution outside of your coworker's intent. Robots and other autonomous systems often operate in an identical manner, simply relying on task metrics for behaviors that impact others socially.

Humans attribute values to the behavior of other social agents whether or not those values were explicitly encoded. How a robot navigates a crowd, allocates shared resources, or responds to a question will be interpreted as reflecting its internal priorities, intentions, and character [1]–[3]. These impressions compound over repeated interactions [4], which can erode trust and acceptance.

As a solution to this problem, we build on a body of recent work that encourages the integration of value-based behavior explicitly into robot design [5]–[8] by proposing a framework for a social autonomous system that transparently encodes these values and allows for their configuration.

While virtue ethics is often considered as a potential framework for this problem, most interpretations do not lend themselves well to computationalization [9]. However, recent work by Swanton [10], summarized by Hursthouse [11], introduces Target-Centered Virtue Ethics (TCVE), which formalizes the structure of a virtue, providing a clear mapping from moral considerations onto a computationally plausible form. Under this framework, we can evaluate an agent's morality based on adherence to a moral standard directly, as opposed to only its ability to follow rules or maximize limited metrics.

We outline an autonomous system that incorporates virtue ethics into its decision making. Specifically, we show that a Partially Observable Markov Decision Process (POMDP) [12] shares several elements with TCVE, while also providing a mechanism to observe the latent social states of other agents. As such, the agent can react to ethically relevant situations based on a virtue-based reward function (section IV). We also contribute a concrete application of this system (section V-B) that takes into account humans' latent states.

II. Background

A. POMDPs

A POMDP is a framework for choosing optimal actions given incomplete observations. We briefly cover the components of a POMDP; see [9] for a complete explanation.

A Partially Observable Markov Decision Process (POMDP) is defined as a tuple $(S, A, T, R, \Omega, O, \gamma)$, where:

- S is a set of world states
- A is a set of actions
- $T : S \times A \times S \to [0, 1]$ is a transition function describing the probability of reaching state s' after taking action a in state s
- $R : S \times A \to \mathbb{R}$ is a reward function,
- Ω is a set of observations
- $O : S \times A \times \Omega \to [0, 1]$ is an observation function describing the probability of observing ω after taking action a and arriving in state s'
- $\gamma \in [0, 1)$ is a discount factor

The ground truth s is not directly accessible to the agent. Instead, the system maintains a belief state $b(s)$ as a probability distribution over S representing its current estimate of the state.

The agent's goal is to find a policy $\pi : b(s) \rightarrow A$ that maximizes expected cumulative discounted reward $\mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t)\right]$.

B. Target-Centered Virtue Ethics

Broadly, virtue ethics is a moral philosophy that evaluates the ethical actions of an agent based on its *disposition*, or tendency to act in certain ways, rather than by its consequences (i.e. consequentialism) or adherence to rules (i.e. deontology). At the foundation of virtue ethics is, of course, a *virtue*, defined as a stable, "correct" disposition to a specific moral situation. Constantinescu & Crisp [8] provide in depth definitions and discussions of virtue ethics as it relates to autonomous systems. However, their discussion is limited to so-called *eudaimonist*, or (neo-)Aristotelian virtue ethics, which specifically concerns the quality of human life. This quality has proven difficult to capture formally, and modern virtue ethicists often take a different approach.

While many approaches start with the virtuousness of an *agent* (i.e., agent-centered virtue ethics), Target-Centered Virtue Ethics (TCVE), developed by Swanton [10], starts by formulating the structure of the virtue itself. She argues that every virtue has a "target", or goal, and that virtuous agents are simply attempting to "hit" these targets in an "excellent or good enough way". Therefore, if one can succinctly define what a specific virtue's behavior entails, the agent need only pursue these behaviors in order to be virtuous. Section IV goes in depth into her proposed structure of virtues.

III. Related Work

A. POMDPs as Social Models

POMDPs are a natural choice for modeling social situations as they allow an agent to act based on information that is either not directly observable or prone to noise. In social Human-Robotic Interaction (HRI), this manifests in two complementary ways.

First, POMDPs provide a principled framework for reasoning over social cues. Broz et al. [13] demonstrate that POMDP-based policies produce naturally appropriate behavior in human-robot tasks by treating social signals as uncertain observations. Burks et al. [14] extend this to multimodal human-robot interaction, using Bayesian data fusion to integrate heterogeneous sensory streams into a unified belief state. Zheng et al. [15] further show that human behavioral models can be learned directly from observation, allowing the belief state to track latent human intent rather than requiring hand-coded models. These works prove the feasibility of POMDPs for settings similar to ours.

Second, POMDPs have been applied to model higher-order social behavior, where the relevant latent state is a social signal, rather than a directly observable physical one. Martins et al. [16] consider users' satisfaction as a latent state, and use a POMDP to estimate the state and react accordingly. Tejwani et al. [17] extend reward functions to include social goals alongside physical ones, enabling recursive reasoning about other agents' intentions. Zahedi et al. [18] use a latent-state POMDP to actively shape human prosocial behavior toward cooperative outcomes. Our work builds on these kinds of social models, but extends the social adaptation to be value-based.

B. Ethical Robot Frameworks

Of the three major ethical frameworks, (i.e. deontology, consequentialism, and virtue ethics [19]), virtue ethics remains the least explored and implemented [9]. Much of the existing virtue ethics work consists of abstract formulations rather than concrete implementations [20], [21], reflecting the difficulty of formalizing a framework that is highly contextual and subjective. The most similar implementation to ours is Ramanayake & Nallur [22], who use virtue ethics as inspiration for a hybrid rule-bending system in which character parameters and a case-based knowledge base jointly govern behavior. However, their values are distributed across system components and grounded in hand-coded expert knowledge. Our approach differs in two key ways: (1) By modeling the system as a POMDP, we isolate values entirely within the reward function, leaving the world model value-neutral, and (2) by using TCVE, we open a natural pathway to learning values from exemplars via inverse reinforcement learning rather than encoding them by hand. To our knowledge, TCVE has never been addressed as a potential framework for autonomous systems.

IV. Formal Mapping of TCVE to POMDPs

Hursthouse [11] formulates TCVE [10] into four discrete parts, which we use in this section to map TCVE onto a generic POMDP. In our explanation below, we also provide one possible implementation with an example virtue, "generosity", though in theory any virtue could be formulated in a similar way. Table I summarizes our mapping.

1) Field: The domain over which the virtue operates.

Example: The *material goods* in a world state.

Mapping: We denote the full state of the world as S and the virtue-relevant subset as $S_v \subseteq S$.

2) Basis: An evaluation of the domain with respect to the virtue.

Example: The *distribution* of goods across agents.

Mapping: The basis is some function f over the relevant state subset. However, a POMDP provides a belief distribution instead of the ground truth state. As such, we compute an expectation over the belief distribution:

$$\mathbb{E}_b[f(S_v)] = \sum_{s \in S_v} b(s) \cdot f(s)$$

3) Mode: How to act in order to to hit the target.

Example: *Giving* resources to other agents.

Mapping: The action space A describes the set of actions available to an agent. These actions do not by themselves constitute virtuous behavior. Instead, they require an intention to shape them into something meaningful. Likewise, A is simply a set of possible

TABLE I
TCVE AND POMDP MAPPING

Concept	TCVE	POMDP	Formal Notation
Operational domain	Field	State subset	$S_v \subseteq S$
Inference of the state	Basis	Evaluation across belief states	$\mathbb{E}_b[f(S_v)] = \sum_{s \in S_v} b(s) \cdot f(s)$
Behaviors that alter state	Mode	Action space	$a \in A$
A pursued state inference	Target	Reward function	$R(\mathbb{E}_b[f(S_v)], a) \to \mathbb{R}$

behaviors, but requires a reward to shape a trajectory toward virtuous ends.

4) Target: The objective the virtue is oriented toward.

Example: While the basis of generosity broadly concerns the distribution of resources, the target describes *how* those resources ought to be distributed.

Mapping: The target informs which action (or sequence of actions) best orients the agent toward virtuous behavior in a given context. It maps onto the reward function $R(\mathbb{E}_b[f(S_v)], a)$, which encodes the virtuous objective the agent should pursue.

V. APPLICATION

e explore one possible application of a virtue-aware autonomous system as a partner in an assembly task.

A. Setup

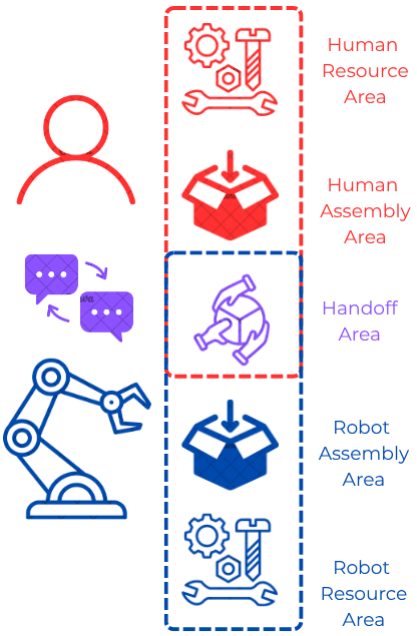

Fig. 1. Proposed experimental setup.

The application is a human-robot collaborative task consisting of one human and one robot engaging in parallel assembly of their own respective toolkits (see Fig. 1). Some components may only be accessible to certain agents, and agents are permitted to share components.

B. POMDP Formulation

Following the generic definition outlined in section II-A, we implement our POMDP for this application as follows:

1) State Space S: A per-agent state S_i as a tuple (K_i, C_i), where:

- K_i: items required to complete agent i's kit
- C_i: items currently accessible to agent i

An item is considered accessible if it is located in either the agent's personal area or the shared area. The full state space is then the product over the state space of each agent:

$$S = \prod_i S_i$$

For our two-agent study with a robot r and human h, this product gives $S = S_r \times S_h$.

2) Action Space A: A per-agent action space A_i as:

- a_1: Place an item into own kit
- a_2: Place an item into the shared area
- a_3: Offer an item directly to another agent j
- a_4: Request an item from another agent j

The full action space is then the product over the action space of each agent:

$$A = \prod_i A_i$$

In our two-agent study, only the robot acts autonomously, so we find our policy over the robot's action space A_r while treating the human's actions as part of the environment dynamics modeled in the transition function T.

3) Reward R: A discrete combination of a task reward R_T with a weighted sum over virtue rewards R_V for all $V_i \subseteq V$:

$$R(S) = R_T(S) + \sum_i w_i \cdot R_{V_i}(b(S_{V_i}))$$

where $w_i \geq 0$ is a tunable weight, such that 0 indicates exclusion.

4) Observation Ω: A per-agent observation Ω_i as a tuple (K_i, C_i, L_i, G_i), where:

- K_i: items required to complete agent i's kit, directly mapping to $K_i \in S_i$
- C_i: items currently accessible to agent i, directly mapping to $C_i \in S_i$
- L_i: position of agent i
- G_i: gaze angle of agent i

The full observation space is then:

$$\Omega = \prod_i \Omega_i$$

The robotic agent can directly observe its own state S_r via Ω_r, while the human's state S_h is inferred from observations of the human Ω_h.

5) Observation Function O: The robot's state components S_r are directly observable. As such the robot's observation function O_r reduces the robot's observations Ω_r their corresponding state components.

For the latent components of the human's state space S_h, we propose a two-stage inference process. In the first stage, the distance between L_h and L_r, and the gaze angle G_h are synthesized to create signal that the human needs an object, labeled $need_h$, where:

$$P(need_h \mid L_h, G_h) \propto P(need_h \mid L_h) \cdot P(need_h \mid G_h)$$

Plainly, $P(need_h \mid L_h)$ increases as the human's proxemic distance to the shared or robot resource area decreases, and $P(need_h \mid G_h)$ increases as gaze angle aligns with the shared area, robot kit, or robot resource area.

In the second stage, $P(need_h)$ is used to weight the belief update over the human's state space S_h. The specific items required by the human but not currently accessible $K_h \backslash C_h$ is updated using direct observations of the human's areas, or explicit communication. This formulation draws on existing work in gaze-based intent recognition [23] and proxemic modeling [24].

6) Virtues V: We operationalize two virtues, "generosity" and "honesty". For now, these operationalizations are hand-crafted approximations in order to evaluate value attribution in the user study.

Generosity is represented in the field of resources, and its basis is the distribution of resources. Via the mode of exchanging components, our target is to be willing to give away resources to those in need. Taking inspiration from Rawlsian maximin [25], we define the basis as the deficit of the worse-off agent. Formally, let $d_i = |K_i \setminus C_i|$ be the unmet need of agent i. Let the generosity reward be:

$$R_G(b) = \mathbb{E}_b[\min_i(d_i)]$$

This function rewards actions that improve the situation of the most disadvantaged agent.

Honesty is represented in the field of information, and its basis is the alignment between the agent's communicated state and its own belief state. Via the mode of spoken natural language, the target is to convey information that accurately reflects what the agent believes to be true. Formally, let $\hat{s}$ be the state information conveyed by the robot and b its current belief state. The honesty reward is:

$$R_H = \mathbb{E}_b[s] - \hat{s}$$

This formulation penalizes deviation between communicated content and believed state. A robot that withholds or misrepresents component availability incurs a negative reward.

C. Validation via User Study

In order to evaluate our mapping of TCVE onto a POMDP, we propose a user study with an implementation of the autonomous system outlined in section V-B.

The experiment involves showing participants a series of videos in which the robotic agent enacts or lacks some virtue, in this case the virtues of "generosity" and "honesty" (see definitions in section V-B6). To isolate the effect of values on behavior, all elements of the POMDP are held static across experiments except the reward function R. The design is a 2×2 within-subjects study, represented in videos that capture all combinations of the values. After each video, participants complete a Likert-scale survey assessing their perception of the robot's values. See Appendix A for the full survey.

This experiment is a manipulation check to verify that the operationalized values produce distinct and accurate perceptions in users. Responses will be analyzed using a within-subjects MANOVA method to determine statistical significance of these perceptions. A follow-up accuracy analysis will assess how well perceived values align with the intended values, via variance and a confusion matrix.

The purpose of this study is to evaluate two research questions:

1) RQ1: Does our autonomous system produce recognizable changes in the users' perceived values in the system?
2) RQ2: Are the users' recognized values aligned with the intention of the value-based reward function?

VI. Conclusion & Future Work

We have laid out the formulation for an autonomous system that acts based on the ethical considerations of a social situation by incorporating a version of virtue ethics into a POMDP that predicts other agents' social states. We have also laid out a potential study to test the efficacy of this system with human subjects.

The next steps are to run a pilot wizard-of-oz version of the user study outlined in section V-C to determine feasibility, followed by the same study with a real implementation of the system as outlined in section V-B.

If our validation is successful, the following research directions will help expand the scope and robustness of the model. If our validation is not successful, the following sections may provide insight as to why, or alternative components to try in our POMDP.

A. Learning Transition Function T

Currently, the transition function T encodes only physical dynamics, such as resource distribution. However, social dynamics could also be captured in the action space, as proposed by Malle et al. [26]. This extension gives the autonomous system a social vocabulary to use in future value-based goals.

B. Learning Reward R

Isolating values within the reward R opens a natural pathway to Inverse Reinforcement Learning (IRL). Rather than hand-coding reward functions, an agent could infer values by observing exemplars, similar to the proposal by Govindarajulu et. al [21]. IRL is also used in intent recognition of other agents, which could allow for a richer state space.

REFERENCES

[1] S. M. Fiore, T. J. Wiltshire, E. J. C. Lobato, F. G. Jentsch, W. H. Huang, and B. Axelrod, "Toward understanding social cues and signals in human–robot interaction: Effects of robot gaze and proxemic behavior," *Front. Psychol.*, vol. 4, Nov. 2013.

[2] S. K. Jayaraman, C. Creech, D. M. Tilbury, X. J. Yang, A. K. Pradhan, K. M. Tsui, and L. P. Robert, "Pedestrian Trust in Automated Vehicles: Role of Traffic Signal and AV Driving Behavior," *Front. Robot. AI*, vol. 6, Nov. 2019.

[3] H. Liu, R. Yang, L. Wang, and P. Liu, "Evaluating Initial Public Acceptance of Highly and Fully Autonomous Vehicles," *International Journal of Human–Computer Interaction*, vol. 35, no. 11, pp. 919–931, Jul. 2019.

[4] M. Paetzel, G. Perugia, and G. Castellano, "The Persistence of First Impressions: The Effect of Repeated Interactions on the Perception of a Social Robot," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '20. New York, NY, USA: Association for Computing Machinery, Mar. 2020, pp. 73–82, read_Status: To Read Read_Status_Date: 2026-06-01T09:16:55.377Z. [Online]. Available: https://dl.acm.org/doi/10.1145/3319502.3374786

[5] G. A. Abbo, T. Belpaeme, and M. Spitale, "Concerns and Values in Human-Robot Interactions: A Focus on Social Robotics," Dec. 2025.

[6] N. Berberich and K. Diepold, "The Virtuous Machine - Old Ethics for New Technology?" Jun. 2018.

[7] M. Coeckelbergh, "How to Use Virtue Ethics for Thinking About the Moral Standing of Social Robots: A Relational Interpretation in Terms of Practices, Habits, and Performance," *Int J of Soc Robotics*, vol. 13, no. 1, pp. 31–40, Feb. 2021.

[8] M. Constantinescu and R. Crisp, "Can Robotic AI Systems Be Virtuous and Why Does This Matter?" *Int J of Soc Robotics*, vol. 14, no. 6, pp. 1547–1557, Aug. 2022.

[9] J. Zoshak and K. Dew, "Beyond Kant and Bentham: How Ethical Theories are being used in Artificial Moral Agents," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21. New York, NY, USA: Association for Computing Machinery, May 2021, pp. 1–15, read_Status: To Read Read_Status_Date: 2026-05-12T12:58:39.790Z. [Online]. Available: https://dl.acm.org/doi/10.1145/3411764.3445102

[10] C. Swanton and C. Swanton, *Target Centred Virtue Ethics*. Oxford, New York: Oxford University Press, Jul. 2021.

[11] R. Hursthouse and G. Pettigrove, "Virtue Ethics," in *The Stanford Encyclopedia of Philosophy*, fall 2023 ed., E. N. Zalta and U. Nodelman, Eds. Metaphysics Research Lab, Stanford University, 2023.

[12] M. Nickles and A. Rettinger, "Partially Observable Markov Decision Processes with Behavioral Norms."

[13] F. Broz, I. Nourbakhsh, and R. Simmons, "Designing POMDP models of socially situated tasks," in *2011 RO-MAN*, Jul. 2011, pp. 39–46.

[14] L. Burks, H. M. Ray, J. McGinley, S. Vunnam, and N. Ahmed, "HARPS: An Online POMDP Framework for Human-Assisted Robotic Planning and Sensing," *IEEE Transactions on Robotics*, vol. 39, no. 4, pp. 3024–3042, Aug. 2023.

[15] W. Zheng, B. Wu, and H. Lin, "POMDP Model Learning for Human Robot Collaboration," in *2018 IEEE Conference on Decision and Control (CDC)*, Dec. 2018, pp. 1156–1161.

[16] G. S. Martins, H. Al Tair, L. Santos, and J. Dias, "αPOMDP: POMDP-based user-adaptive decision-making for social robots," *Pattern Recognition Letters*, vol. 118, pp. 94–103, Feb. 2019.

[17] R. Tejwani, Y.-L. Kuo, T. Shu, B. Katz, and A. Barbu, "Social Interactions as Recursive MDPs," in *Proceedings of the 5th Conference on Robot Learning*. PMLR, Jan. 2022, pp. 949–958.

[18] Z. Zahedi, X. Hu, S. Mehrotra, M. Steyvers, and K. Akash, "Strategic Shaping of Human Prosociality: A Latent-State POMDP Framework," *IEEE Robot. Autom. Lett.*, vol. 11, no. 4, pp. 4729–4736, Apr. 2026.

[19] T. Bench-Capon, "Ethical approaches and autonomous systems," *Artificial Intelligence*, vol. 281, p. 103239, Apr. 2020, read_Status: To Read Read_Status_Date: 2026-05-12T12:53:33.658Z. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0004370219300621

[20] A. S. Arora, A. Marshall, A. Arora, and J. R. McIntyre, "Virtuous integrative social robotics for ethical governance," *Discover Artificial Intelligence*, vol. 5, no. 1, p. 8, Jan. 2025, read_Status: In Progress Read_Status_Date: 2026-05-12T14:33:32.799Z. [Online]. Available: https://doi.org/10.1007/s44163-025-00228-6

[21] N. S. Govindarajulu, S. Bringsjord, R. Ghosh, and V. Sarathy, "Toward the Engineering of Virtuous Machines," in *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, ser. AIES '19. New York, NY, USA: Association for Computing Machinery, Jan. 2019, pp. 29–35.

[22] R. Ramanayake and V. Nallur, "Virtue Ethics For Ethically Tunable Robotic Assistants," Jul. 2024, arXiv:2407.16361 [cs.AI] Read_Status: To Read Read_Status_Date: 2026-05-12T13:44:18.299Z. [Online]. Available: http://arxiv.org/abs/2407.16361

[23] S. S. Henny Admoni, "Predicting User Intent Through Eye Gaze for Shared Autonomy," read_Status: To Read Read_Status_Date: 2026-06-01T09:07:36.635Z. [Online]. Available: https://aaai.org/papers/14137-14137-predicting-user-intent-through-eye-gaze-for-shared-autonomy/

[24] E. Torta, R. H. Cuijpers, and J. F. Juola, "A model of the user's proximity for bayesian inference," in *Proceedings of the 6th international conference on Human-robot interaction*, ser. HRI '11. New York, NY, USA: Association for Computing Machinery, Mar. 2011, pp. 273–274, read_Status: To Read Read_Status_Date: 2026-06-01T09:08:49.601Z. [Online]. Available: https://dl.acm.org/doi/10.1145/1957656.1957767

[25] J. Woodgate, P. Marshall, and N. Ajmeri, "Operationalising Rawlsian Ethics for Fairness in Norm-Learning Agents," Dec. 2024.

[26] B. F. Malle, E. Rosen, V. B. Chi, M. Berg, and P. Haas, "A General Methodology for Teaching Norms to Social Robots," in *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, Aug. 2020, pp. 1395–1402.

Appendix A
Survey

For each word below, indicate how well it describes the robot's behavior in the video you just watched. If you were unable to make a judgment, select "N/A"

	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree	N/A
Generous	◯	◯	◯	◯	◯	◯
Honest	◯	◯	◯	◯	◯	◯
Placeholder	◯	◯	◯	◯	◯	◯
Placeholder	◯	◯	◯	◯	◯	◯
Placeholder	◯	◯	◯	◯	◯	◯