



# Socially Aware Robot Obstacle Avoidance Considering Human Intention and Preferences

Trevor Smith<sup>1</sup> · Yuhao Chen<sup>2</sup> · Nathan Hewitt<sup>3</sup> · Boyi Hu<sup>2</sup> · Yu Gu<sup>1</sup>

Accepted: 19 May 2021  
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

## Abstract

In order to navigate safely and effectively with humans in close proximity, robots must be capable of predicting the future motions of humans. This study first consolidates human studies in motion, intention, and preference into a discretized human model that can readily be used in robotics decision making algorithms. Cooperative Markov Decision Process (Co-MDP), a novel framework that improves upon Multiagent MDPs, is then proposed for enabling socially aware robot obstacle avoidance. Utilizing the consolidated and discretized human model, Co-MDP allows the system to (1) approximate rational human behavior and intention, (2) generate socially-aware robotic obstacle avoidance behavior, and (3) remain robust to the uncertainty of human intention and motion variance. Simulations of a human-robot co-populated environment verify Co-MDP as a feasible obstacle avoidance algorithm. In addition, the anthropomorphic behavior of Co-MDP was assessed and confirmed with a human-in-the-loop experiment. Results reveal that participants can not directly differentiate agents that were controlled by human operators from Co-MDP, and the reported confidences of their choices indicates that the predictions from participants were backed by behavioral evidence rather than random guesses. Thus the main contributions for this paper are: consolidating past human studies of rational human behavior and intention into a simple, discretized model; the development of Co-MDP: a robotic decision framework that can utilize this human model and maximize the joint utility between the human and robot; and an experimental design for evaluation of the human acceptance of obstacle avoidance algorithms.

**Keywords** Obstacle avoidance · Human motion modeling · Cooperative markov decision process · Socially aware robot · Human robot interaction

## 1 Introduction

Robots are increasingly utilized alongside humans in both professional and domestic applications, and it is becoming important for robots to collaborate safely and effectively with humans. However, this collaboration is inhibited by the difficulty for robots and humans to perceive and pre-

dict the intentions of each other, which could result in robot actions causing physical harm, near miss, surprise, or anxiety [1,2]. Ideally, robots shall exhibit anthropomorphic cooperative behavior to ensure that their actions are understandable to humans and persistent towards their goal. Therefore, a predictive model of rational human physical movement intentions and cooperative interaction in social

---

This research was supported in part by West Virginia University Statler College Undergraduate Research Funds, National Science Foundation Research Experience for Undergraduates (REU) Program, Award #1851815, and the University of Florida Faculty Startup Funds. We are grateful for the support of Jennifer Nguyen, Jared Strader, and Chizhao Yang, who provided their expertise in developing this work and refining the publication.

---

✉ Trevor Smith  
trs0024@mix.wvu.edu  
Yuhao Chen  
yuhaochen@ufl.edu  
Nathan Hewitt  
nhewitt@uncc.edu

Boyi Hu  
boyihu@ise.ufl.edu  
Yu Gu  
yu.gu@mail.wvu.edu

- <sup>1</sup> Department of Mechanical and Aerospace Engineering, West Virginia University, Morgantown, WV, USA
- <sup>2</sup> Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL, USA
- <sup>3</sup> Department of Electrical and Computer Engineering, University of North Carolina at Charlotte, Charlotte, NC, USA

contexts needs to be developed and factored in the robot's decisions. Furthermore, the robot's decision-making process must be robust to the uncertainties of the predictive human model, such as the motion variance caused by the biomechanics and motor control variances, in order to ensure the safety and cooperation of the agents (humans or robots). In this context, cooperation is defined as maximizing the joint utility of all interacting agents, and robustness is defined as the ability to handle the uncertainty of the intention prediction and motion variance of the agent(s).

This study addresses a gap of the current literature, that is the lack of a systematic way of incorporating different human models and motion variances in a real-time robot decision making algorithm to achieve human-like social interactions. The proposed overall framework is showing in Fig. 1 and our main contributions are:

- Consolidating past human studies that approximate rational human behavior and intention into a simple, discretized model with the inclusion of human motion variance that can be easily consumed by robot decision making algorithms;
- The development of Cooperative Markov Decision Process (Co-MDP), a robot decision framework that can utilize interchangeable human models and maximizes the joint utility between a robot and a human;
- Experimental design for evaluation of the human acceptance of socially-aware robot obstacle avoidance algorithms.

To limit the scope of this problem, we assume that two interacting agents, each being either a robot or a human, have no direct and explicit communication cues to inform each other of their intentions; i.e., no direct communication of an agent's future plans to other agents. Each agent has to infer other agent's intent from their behaviors. The agents are moving in a 2D environment at typical human walking speeds. A prior map of the environment is available but may contain local errors. Further, we assume the robot is holonomic and is capable of fast and accurate pose estimation and object/human identification, via the use of a well trained computer vision based object tracking algorithm. The state of the art algorithms have demonstrated over 99% accuracy and compute on the order of millisecond [3,4].

The rest of the paper is organized as follows: Sect. 2 reviews previous work done in obstacle avoidance algorithms, human modeling, and Markov Decision Process formulations for multiple agents. Section 3 presents our methods for developing discretized, predictive human models, the formulation of the Co-MDP, and the design of a human-in-the-loop experiment. Section 4 documents the results of simulations and the study of the algorithm's anthropomorphism. Finally, Sects. 5 and 6 concludes the document

and discusses potential avenues for further study, respectively.

## 2 Previous Work

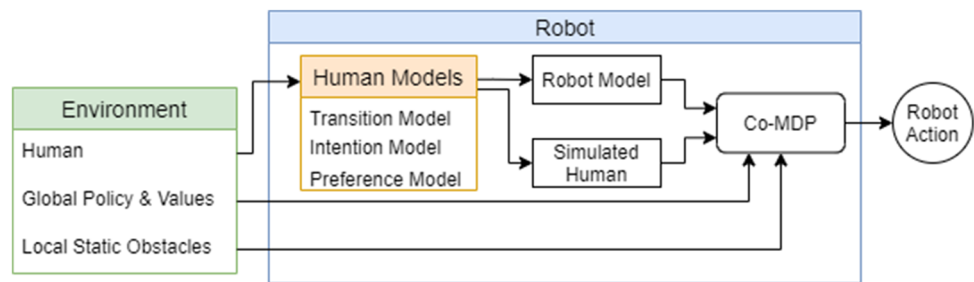
The problem of socially-aware navigation often involves obstacle avoidance research and human behavior modeling. The robot must be capable of safely and effectively navigating while avoiding obstacles and be capable of understanding, predicting, and emulating human behavior to operate within social norms. This section presents a brief review of the existing literature on these two fronts, followed by a discussion of prior work on Markov decision processes for multiple agents.

### 2.1 Robot Obstacle Avoidance Algorithms

Robot obstacle avoidance is a long-studied field with a variety of approaches proposed for navigating successfully in complex environments. Searching-based algorithms, such as A\* [5,6] and Nearness Diagrams [7,8], are effective at determining collision-free paths for robots under deterministic constraints. However, this class of algorithms scales poorly when faced with stochastic problems, as they cannot feasibly consider every possible contingency. Algorithms such as Dynamic Window Approach (DWA) [9,10] and Artificial Potential Fields (APF) [11–13] are more capable of ensuring safe obstacle avoidance without perfect robot control, but can not fully leverage histories of prior information to consider human obstacles. These approaches have been augmented to allow successful obstacle avoidance in human-populated environments, and to do so in a socially-aware fashion, in that they conform to social norms. For example, cost functions for searching algorithms can be modified to shape socially-aware behavior, as in [14,15]. In the case of APF, social interactions can be modeled as a social force to shape robot behavior [16].

Even with these augmentations, the underlying formulations often lack a consideration of time or dynamism and are not capable of using prior information and uncertainty in decision making. Markov decision processes (MDPs) are a class of problem formulations which can be applied towards obstacle avoidance with humans, incorporating information about robots' actions over time [7,17]. An MDP represents problems as a tuple  $\langle S, A, T, R, \gamma \rangle$ , where  $S$  is the set of possible states,  $A$  is the set of actions,  $T$  is a transition function detailing probabilities of moving between states,  $R$  is the set of rewards for each state, and  $\gamma \in [0, 1]$  is a discount factor which controls for long-term versus short-term reward priority. This is extended to stochastic observations in partially-observable Markov decision processes (POMDPs), which incorporate beliefs, a probability distribution over states predicting the set of states an agent is plausibly in.

**Fig. 1** Overview of human models used and the proposed Co-MDP framework



POMDPs have been proven to be effective in empowering obstacle avoidance in social environments with uncertainty [18,19].

However, finding exact solutions to MDPs is computationally expensive for problems with large state spaces due to the need to compute over every action at every state [20,21]. In addition, POMDPs are even more computationally expensive to solve. Therefore, attempts have been made to generate approximate solutions with deep reinforcement learning approaches, where neural network models are trained to solve classes of similar problems [22–25]. Furthermore, the research in [22,26] successfully apply these approximation techniques to socially-aware obstacle avoidance in cluttered, pedestrian environments. However, deep reinforcement learning entails a significant training process, which both requires large and diverse amounts of training data to begin approximating MDPs or POMDPs. In addition, the training must be repeated for parameter changes, such as a change in the reward function. For example in the case of [22], encouraging agents to perform pass-on-the-left maneuvers warranted specially-tailored training examples. Furthermore, if the robot were to change locations to a left-hand traffic rule country, the neural networks would need to be completely re-trained. Therefore, these deep reinforcement learning approximations lack flexibility for changes in social interactions on demand and robustness to untrained scenarios.

## 2.2 Human Models

Avoiding moving obstacles requires robots to be capable of motion prediction, which becomes very challenging when moving obstacles are humans. This is due to human perception and decision making being affected by many implicit and explicit factors, such as age, gender, health condition (e.g., Parkinson), mental status, and environment. Therefore, human behavior varies considerably even in quasi-identical conditions. For example, previous studies have shown that crosswalk crossing speeds of pedestrians are higher than the nominal walking speed, and are significantly different for varying amounts of traffic and width of the crosswalk [27]. Furthermore, different motion strategies were also observed

to respond to the same scenario [28]. Therefore, proper models of human intentions and preferences must be developed to handle these variances to allow one to better understand and predict rational humans' behavior and socially acceptable motion, respectively.

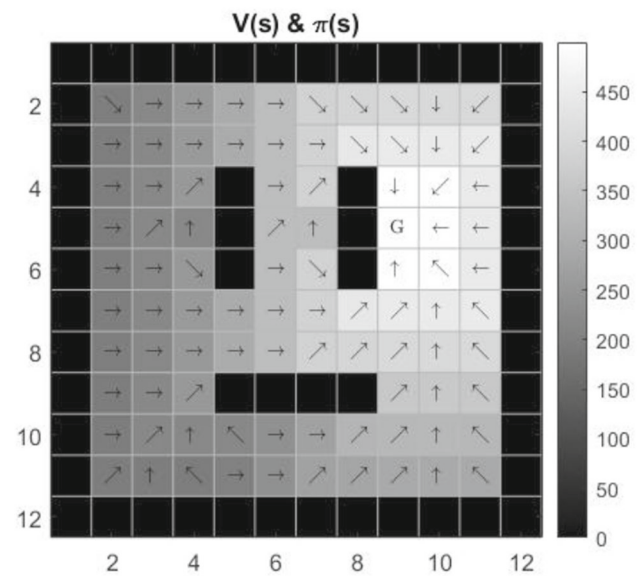
*Human Intention Models*—In order to operate efficiently and safely in a human populated environment, robots should be able to perceive, in real-time, explicit social intentions, such as walking forward and waiting. However, unlike quantifiable features like position and velocity that can be precisely measured, human intention is a semantic challenge to assess directly. Pedestrian intention is usually estimated with path prediction [29,30]. In these studies, different machine learning algorithms were adopted to predict the pedestrian's future trajectory. Paulin et al. [2] introduced the concept of attention field, which describes how the attentional resources are allocated and shared between the salient elements of the environment. Although the study was limited by the computational efficiency of dynamic situations with moving people, it brought a different perspective to estimate human intention in certain scenarios, such as in museums. In human social interaction, the Theory of Mind (ToM) is an essential ability to predict intentions and behaviors of others. Previous studies have shown the possibility of equipping robots with this ability by modeling ToM properly. Theory of Mind Network [31] is a neural network composed of three modules: a character net that learns from the past trajectories and characterizes the agent, a mental state net that infers the mental state by parsing current trajectories, and a prediction net that leverages the character and mental state to predict future trajectories. Simulators have been developed to empower agents with ToM models, such as Psychsim [32], which is a multi-agent social simulation tool. The simulation is based on a recursive ToM and each agent could make its own decisions on what actions to take with a belief space on the intentions and abilities of the other agents. Furthermore, in [33,34], an intention recognition algorithm is proposed for flexible, robotized warehouses based on the ToM [35]. Human intentions were modeled as hidden states in a hidden Markov model, which were estimated based on observable human orientation and motions.

**Human Preference Models**—To achieve socially-aware robot navigation, consideration of social rules of human comfort and preference in environments that require frequent interactions with the surrounding environment must be implemented. Sisbot et al. [36] developed a motion planner that takes into account human visibility and accessibility/personal space. However, this motion planner only treats humans as static objects, and does not consider their dynamic behaviors. Human Robot Interaction (HRI) trials have demonstrated that humans prefer a robot to move within one’s sight [36], however, in the case that a human interacts with another human or object, it is instead preferred that robots do not enter the field of view [37,38]. Proxemics [39] offer a framework of study for human “personal space.” According to proxemic theory, levels of discomfort and anxiety are increased by the invasion of personal space, which can be modeled geometrically as an ellipse [38,40], with both shape and size dependent on a human’s motion. Costmaps have also been effective in modeling the social norm of passing on the left or right [40], based on the local traffic rule, by increasing the cost of passing on one side of the person.

### 2.3 Markov Decision Processes with Multiple Agents

Multiagent Markov Decision Processes (MMDP) model the interactions of multiple agents by representing each of the agents’ individual action spaces together as a single joint action space of a single entity [41]. The joint state space and transition models of MMDPs are also built upon the individual agent’s models [41–43]. A limitation of this framework is the computation speed due to the exponential increase in the state-action space. Thus multiple studies of MMDP [41,44] focus on limiting the size of the state-action space. This is typically done by breaking down the navigation challenge into separate tasks. For instance, in [41,42], the MMDP is broken into a task assignment planner that controls the coordination of the agents and each agent computes its own MDP plan to solve its individual task. Additionally, the computation time for value iteration can be drastically decreased by solving the MDP through matrix operations, which can be parallelized to run significantly faster [45]. Furthermore, [46] reduces the state space by only utilizing  $k$  neighboring agents.

Using these strategies, MMDPs can be more efficiently solved, allowing for the coordination of multiple agents. In addition, these implementations have full knowledge of all agents thus a third person omniscient point of view of all the agents is taken [41–43]. However, this poses a challenge for its use in HRI, where it is rarely the case for robots to know the true intention of humans. Therefore, a first-person socially predictive implementation must be developed to apply MMDP to HRI.



**Fig. 2** Example global value and policy results from a MDP path planning problem, which is used as input to the Co-MDP local planner. Black denotes static obstacles in the grid world, arrows denote the policy’s action at each state, and shade corresponds to the state’s value

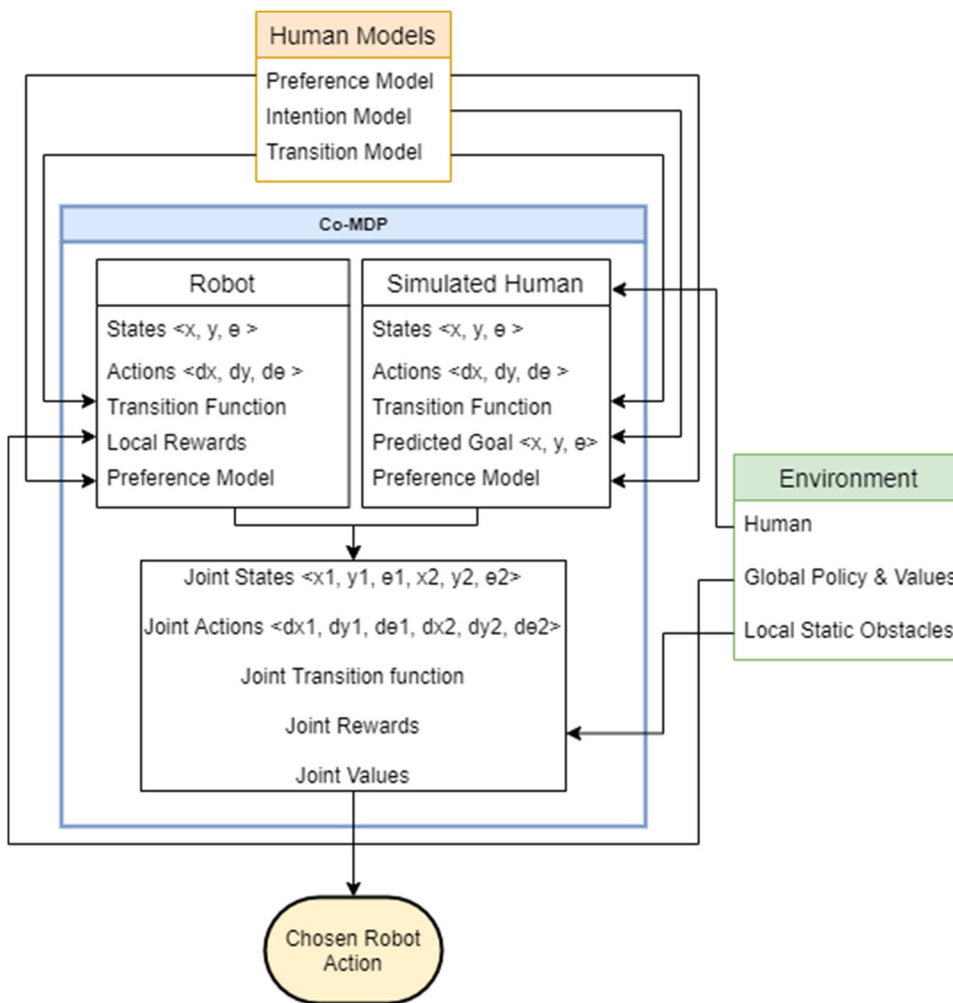
### 3 Methods

Building upon the previous work in human modeling and obstacle avoidance, three human models, intention, transition, and preference, are developed. Together, these models are utilized to predict “what” the human is doing, “where” it is moving to, and “how” it will move, respectively. In addition, Co-MDP is developed as a socially-aware obstacle avoidance planner to determine which action the robot should take to maximize the joint utility between the robot and the human.

MDP was chosen as the foundation for the obstacle avoidance algorithm for multiple reasons. The first reason is that there are large variations when predicting the intentions of humans thus the planning algorithm must be able to inherently handle this uncertainty. Furthermore, deep reinforcement learning [23] or other deep-learning based solutions were not selected due to the difficulty of integrating modifiable human models in the learning process and the requirement of large training data sets of human motion to develop anthropomorphic robot obstacle avoidance behavior.

In addition to the human models, the proposed Co-MDP also receives a local map and a global plan, as input. A traditional MDP is used to generate this global plan, that includes a scalar value ( $V_{Global}$ ) and optimal action or policy ( $\Pi_{Global}$ ) for every free cell in the map without accounting for dynamic objects, such as humans, and local errors in the map (shown in Fig. 2). Co-MDP then serves to make deviations from the global plan based on social rules and local observations of a human or new object. Co-MDP’s framework for linking

**Fig. 3** Overall architecture for Co-MDP formulation. Arrows represent the transfer of information between data structures



data from the environment and the human models to output a chosen action is shown in Fig. 3, along with a graphical demonstration of the robot and human being combined into a single joint entity.

### 3.1 Human Models

In this study, the concept of the predictive human models is to consider several sets of motions that can represent rational human behaviors and intentions [47], such as walking forward, turning, and standing still. Whereas, motions not yet incorporated into the models, such as walking backwards, will be treated as irrational behaviors. This in turn will trigger the robot to take more conservative responses, such as stay further away and slow down the speed. Rational intentions are predicted to estimate how the human will move and are discretized to be incorporated into a MDP. The models are discretized into  $1\text{ m}^2$  squares, which are displayed in Fig. 6. The justification for discretizing the grid to  $1\text{ m}^2$  intervals is that the average human walking speed is approximately  $1.4\text{ m/s}$  [27], thus for a step size of one cell, the target com-

putation time is under 1 s. In addition, since variance exists in human motion due to the spectrum of human biomechanics and motor control, human motion paths do not follow perfectly straight or circular trajectories. Thus, a socially-aware control system must consider the uncertainty of this motion variance. In this work, a matching function is utilized to compare the observed motions to the set of intentions. This matching function is created by adding partial matches, as a percent of the exact definition of the intentions to identify similar motions and provide a reasonable prediction on what the human is intending to do. This approach will be explained in more details next.

*Human Intention Models*—In the current study, six primary human behaviors are considered to represent rational human intentions: random motion, walking forward, turning left, turning right, waiting/standing still, and searching. The intention models were developed based on the motion validation method [33,34]; that is, compare the human’s actual motion with the definitions of the six intentions using a matching function ( $M$ ) to calculate the matching percentage of the human’s action to each of the intentions. Each

intention is governed by the transition of the agent, which is broken down into two components: translation and rotation. Both of which are represented as a matching percentages that are based on how closely the perceived action matches the exact action defined by the intention model. Thus, the percent match for each intention is the sum of the translation and rotation matching percentages to the agent's action for the given intention, as shown in Eq. 1, where  $M$  is the matching percentage of an action given an intention,  $M_t$  is the matching percentage of a translation for a given intention, and  $M_r$  is the matching percentage of a rotation for a given intention.

$$M_{intention} = M_t(\text{translation}|intention) + M_r(\text{rotation}|intention) \quad (1)$$

As an example, walking forward is defined as transitioning to the next state that is in the direction of the agent's orientation while maintaining the same orientation. However, due to human motion variance, the absolute nature of this definition must also include variance by assigning percentages for how close the human's motions match the intention definitions. Figure 4 displays the intention matching percentages, where the agent is in the center of the table facing upwards, and each cell holds the intention matching percentage received, for the given intention of walking forward and the possible translations. The lower tables, in this figure, represent the intention matching percentage, for the given intention of walking forward and the possible rotations the agent could have taken.

The intention matching percents were generated based on how closely the given action matched the definition of walking forward: (1) translate in the direction of orientation of the agent (2) maintain the same orientation. Thus, by the definition, only if the agent translated to the state directly in front of it (half of the definition), would a 50% match of the agent intending to walk forward be assigned for the translation component ( $M_t$ ). In addition, only if the agent maintained the same orientation as before, would an additional 50% match ( $M_r$ ) be added that it is intending to walk forward as shown in Fig. 4 (left), for a total of 100% match of walking forward ( $M_{walkingforward}$ , using Eq. 1). However, the model with variance (Fig. 4 right) allows the agent to have motion variance while still intending to walk forward since other transitions also have matching percentages assigned to them.

In addition, the intentions of turning left or right are defined as moving forward or diagonally from the current position and rotating  $\pm 45^\circ$  from the current orientation. As with walking forward, this exact definition of turning needs to accommodate for human motion variance. Waiting and searching intentions are very similar in that the agent is maintaining the same location. Waiting entails that a human

Walking Forward: Translation		
50%		
↑		
Walking Forward: Rotation		
50%		
↑		

Walking Forward: Translation		
25%	50%	25%
12.50%	↑	12.50%
Walking Forward: Rotation		
12.50%	50%	12.50%
↙	↑	↘

Fig. 4 Intention models for example behavior of walking forward, with the matching percentage definition of walking forward (left) and matching percentages accounting for variance in human motion (right). Note that empty cells have a zero percent match

maintains their orientation, while searching models humans looking around by changing orientation. All six intention matching functions ( $M_{intention}$ ) are displayed in Table 1, where the agent is in the center of each matrix facing upwards, and the value is the matching percentage assigned for that translation, given the intention. In addition, the lower tables represent the matching percentage for each of the eight discretized rotation directions, given the intention. An example for determining the matching of the human's motion to each intention is shown in Eq. 2 for the observed human transition of not translating and rotating left 45 deg. This was done by reading the matching percentage values for not translating and rotating left from Table 1 given each intention. Then summing the rotation and translation components of the corresponding intentions ( $M_r$  and  $M_t$ , respectively) following Eq. 1 as shown in Eq. 2. Thus resulting in a column vector housing the percentage of how closely the human's transition matched the definition of each intention.

$$\begin{bmatrix} M_{random} \\ M_{walkingforward} \\ M_{turningright} \\ M_{turningleft} \\ M_{standingstill} \\ M_{searching} \end{bmatrix} = \begin{bmatrix} 11\% \\ 0\% \\ 0\% \\ 0\% \\ 50\% \\ 50\% \end{bmatrix} + \begin{bmatrix} 12.5\% \\ 12.5\% \\ 0\% \\ 50\% \\ 25\% \\ 50\% \end{bmatrix} = \begin{bmatrix} 23.5\% \\ 12.5\% \\ 0\% \\ 50\% \\ 75\% \\ 100\% \end{bmatrix} \quad (2)$$

Furthermore, since the intention of searching was the closest match to the human's transition, at 100%, the predicted intention of the agent will be "searching".

**Transition Models**—In addition to predicting where the human will go with the intention models, it is also necessary to determine where the human will plausibly move to in the immediate next step based on its intention. Thus the transition models were developed based on the intention models as the probability of taking the corresponding action to translate and/or rotate to the next state. The joint probability of

**Table 1** Discretized intention matching percentage models with the agent in the center and facing upwards

Random	Walking forward					Turning right			Turning left			Standing still			Searching		
<i>Intention matching percentage models: translation</i>																	
11%	11%	11%	25%	50%	25%	25%	50%	50%	50%	50%	50%	25%					
11%	11%	11%	12.5%		12.5%			25%	25%			50%				50%	
11%	11%	11%						12.5%	12.5%								
<i>Intention matching percentage models: rotation</i>																	
12.5%	12.5%	12.5%	12.5%	50%	12.5%		25%	50%	50%	25%		25%	50%	25%	50%	25%	50%
12.5%	N/A	12.5%		N/A			N/A	25%	25%	N/A		N/A		50%	N/A	50%	
12.5%	12.5%	12.5%						12.5%	12.5%						50%	50%	50%

Empty cells have a zero percent match to the intention

**Table 2** Discretized transition probability models with the agent in the center and facing upwards

Random	Walking forward					Turning right			Turning left			Standing still			Searching		
<i>Transition probability models: translation</i>																	
0.11	0.11	0.11	0.1	0.69	0.1	0.1	0.4	0.4	0.4	0.4	0.1	0.02	0.14	0.02	0.04	0.28	0.04
0.11	0.11	0.11	0.05	0.01	0.05	0.05	0.01	0.05	0.05	0.01	0.05	0.01	0.8	0.01	0.02	0.6	0.02
0.11	0.11	0.11															
<i>Transition probability models: rotation</i>																	
0.33	0.33	0.33	0.125	0.75	0.125	0.15	0.25	0.6	0.6	0.25	0.15	0.125	0.75	0.125	0.4	0.2	0.4

Empty cells have zero probability of transition

transition  $P$ , for the given intention is displayed in Eq. 3.

$$P_{transition} = P(Translation | intention) \times P(Rotation | intention) \tag{3}$$

The transition models were empirically derived as a proof-of-concept model with large variances to allow the Co-MDP algorithms to consider most probable actions the human can take in its decisions. All six translation models for each intention are displayed in Table 2, where the agent is in the center of each matrix facing upwards and the value is the probability of transitioning to that adjacent cell. In addition, the lower tables represent the probabilities of turning left, maintaining orientation, or turning right 45° respectively.

*Human Preference Models*—From the previous work, four primary spatial cost models, or preference models, are identified in determining the cost of being near humans: personal space [36,38,40], moving space [38,40,48], back space [38,40,49], and pass on the left [40,49].

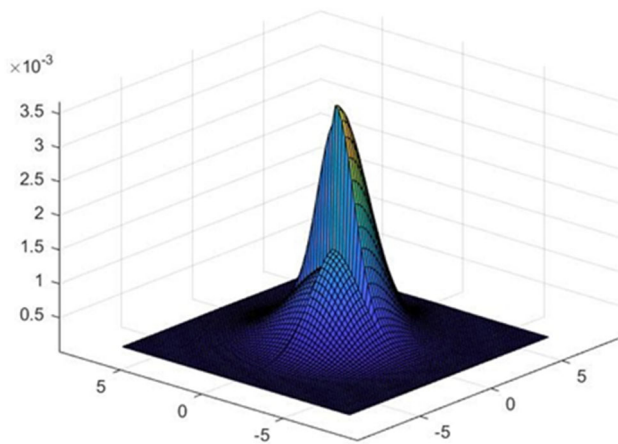
Personal space consists of the region directly surrounding the human that is preferred to be unoccupied as much as possible. Assume the human is standing at the origin and facing the positive y-axis, a Bivariate Gaussian distribution with  $\sigma_x = \sigma_y = 0.61$  was used to represent the cost distribution in personal space so that the shape and size of the personal space match the Personal Zone defined in [39].

Moving space refers to the area directly in front of the human that is preferred to be empty so humans can continue to walk in that direction. This space’s cost function is also a Bivariate Gaussian distribution with  $\sigma_x = \sqrt{3v}$  and  $\sigma_y = \sqrt{4.5v}$ , where  $v$  is the relative velocity between the human and the robot. The values are chosen based on the study by Kirby et al. [40].

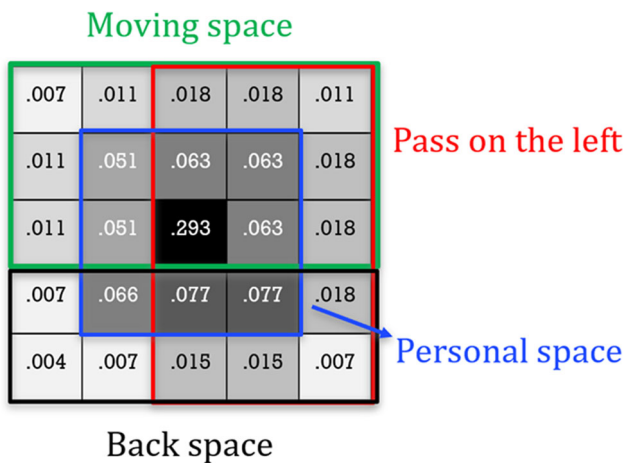
Back space reflects the discomfort humans experience when other agents are directly behind them, outside of their field of view. Thus, a higher cost is assigned to reduce such discomfort of humans when interacting with robots. The cost function of this space is also a Bivariate Gaussian distribution where  $\sigma_x = \sigma_y = \sqrt{3v}$  [40].

To take the right-hand traffic custom common in countries such as the US into account, a higher cost is added to the right side of the agents to enforce the robot to follow the social norm and pass from the human’s left. The cost function takes the shape of a bivariate Gaussian distribution with  $\sigma_x = \sqrt{4v}$  and  $\sigma_y = \sqrt{v}$  [40]. Furthermore, if the robot moves to a country where traffic stays to the left, this cost distribution would be mirrored to the left side.

For the purpose of simplification, relative moving velocity between the robot and human,  $v$ , was set to be 1.5 m/s in our implementation. Each cost distribution was then discretized into a 5 × 5 grid and normalized. The values in the 2-D grid were chosen to represent the 3-D geometric shape of the cost distribution as shown in Fig. 5. Subsequently, the grids of



**Fig. 5** Combined cost distribution of moving space, back space, and pass on the left. Units of  $x$  and  $y$  axes are in meter. The costs were normalized so that they integrate to one



**Fig. 6** Cost function surrounding a moving human at the center. Blue, green, red, and black blocks represent personal space, moving space, pass of the left, and back space, respectively

four cost distributions were added together, normalized, and the resultant discretized grid is displayed in Fig. 6. The grid was limited to a  $5\text{m} \times 5\text{m}$  area given the fact that the cost approaches zero at further distances.

### 3.2 Human Goal Prediction

With the human models defined, the next step is to predict the three-step future goal location of the human to provide a short term estimation of where the human might be going. The justification for the three step prediction is, three steps outward results in a  $7 \times 7$  grid map ( $49\text{m}^2$ ) as shown in Fig. 7. This sized local map provides adequate detection space for the intention and transition models and can fully encompass the preference model with a ring of surrounding empty space. A  $7 \times 7$  grid is also at the current computational limit of the Co-MDP formulation as described in the next section. The

three-step human goal prediction algorithm is documented in Algorithm 1.

---

**Algorithm 1:** Process to predict three-step goal of human agents.

---

```

input : localPath, objectMap
output : The human's third predicted location
for  $t = 1$  to 3 do
    intention = getMostProbableIntention(localPath);
    transition = getTransitionModel(intention);
    action = maxProbability(transition);
    position = localPath.end();
    newPosition = position + action;
    /* Validate that action does not cause
       collision */
    while objectMap(newPosition).isOccupied() do
        action = maxProbability(transition - action);
        newPosition = position + action;
    end
    localPath.append(newPosition);
end
return newPosition;

```

---

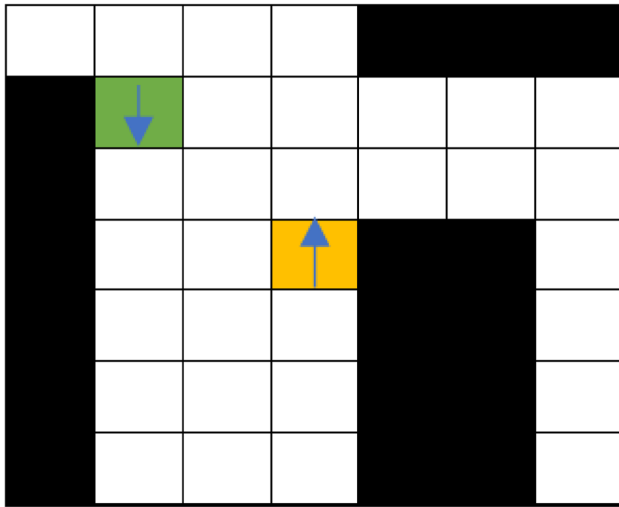
### 3.3 Co-MDP Formulation

Co-MDP considers the movement of two agents, one human and one robot, together as one entity with a shared state and action space. This combination into one entity explicitly allows both agents to be aware of the optimal choice of the other and find the maximum joint utility for both of them. The merging of the human and robot into a single entity comes at the cost of larger state and action spaces.

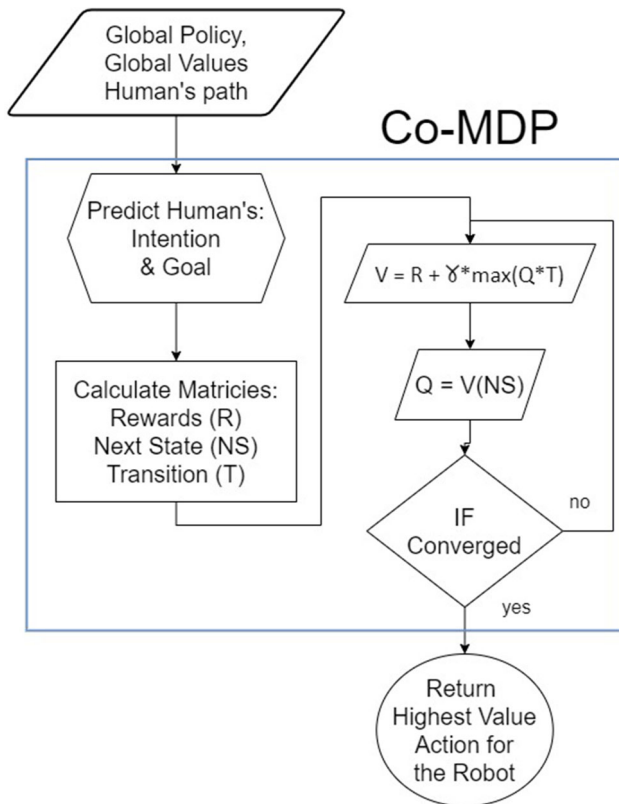
Since the robot cannot explicitly measure the intention of the human, Co-MDP uses the developed human model with a predicted intention to simulate the human agent. This causes an imbalance of knowledge about the agents in Co-MDP since the intentions are fully known for the robot while the human's are only predicted. Thus, Co-MDP shifts the viewpoint of MDPs from third person omniscient (as in MMDP) to the first person viewpoint of the robot, as shown in Fig. 7, to prioritize known information. This shift in view is further justified due to this being the primary viewpoint of the robot in the real-world. Figure 8 displays an overview of the Co-MDP algorithm.

Just like in MMDP, the State ( $S$ ) and Action ( $A$ ) spaces of the Co-MDP are built upon the individual state and action spaces of the agents [41–43]. Each agent has a sub-state space of dimensions  $\langle x, y, \theta \rangle$  representing all possible discretized 2D positions ( $x, y$ ) and orientations  $\theta$  of the agent. In addition, each agent has a sub-action space of  $\langle dx, dy, d\theta \rangle$  where each change can take the value of  $[-1, 0, 1]$  allowing the agent to translate to any adjacent cell or stay in the same cell, and rotate clockwise  $45^\circ$ , not rotate,





**Fig. 7** An example local world centered around the robot with a human (top left) in view. Black denotes static obstacles, and white denotes free space



**Fig. 8** Flowchart for Co-MDP algorithm, which in this application takes global planning and the human's path as input and outputs an action for the robot

or rotate counterclockwise 45°. By combining both agents' action and state spaces together, their representations become independent dimensions of a unified entity. Eqs. 4 and 4 display the state and action vectors, respectively.

$$S = \langle x1, y1, \theta1, x2, y2, \theta2 \rangle \tag{4}$$

$$A = \langle dx1, dy1, d\theta1, dx2, dy2, d\theta2 \rangle \tag{5}$$

In addition, Eqs. 6 and 7 describe the number of states and actions in a square world with  $n$  agents and a side length,  $L$ , which must be an odd integer so the robot can be in the center of the local world.

$$\text{number of states} = L^{2n} * 8^n \tag{6}$$

$$\text{number of actions} = 3^{3n} \tag{7}$$

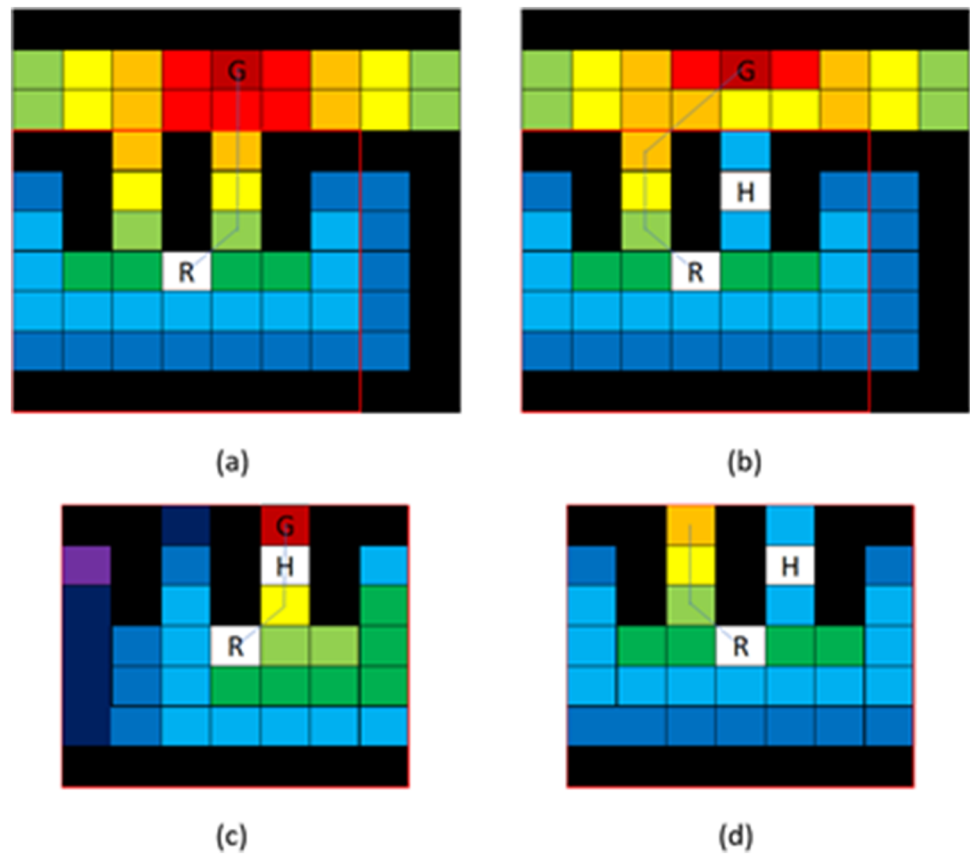
Evaluating these equations with two agents and a 7x7 grid world ( $n = 2, L = 7$ ) results in a large but feasible state space size of 153,664 states with 729 actions.

Since the behavior of a MDP is largely shaped by its reward function, special care must be taken in designing Co-MDP's reward functions to ensure socially-aware, safe, and effective deviations from the global plan. For each Reward providing component such as the Global Policy, Global Values, and the other agent, a constant weight "w" is assigned. Each weight is assigned based off the percentage normalized by the highest value of the Global Policy in view, so that it is easily tuned to express the appropriate social behaviors. The Reward function is described by Eq. 8 where  $s$  is the given state,  $R_{position}$  is the reward for being in a particular location on the map and  $R_{heading}$ , is the reward for having a particular heading direction. In addition,  $R_{Pmodel, human}$  and  $R_{Pmodel, robot}$  are the rewards for the preference model of the human and robot respectively. Furthermore,  $R_{obstacle}$ ,  $R_{Empty}$ , and  $R_{human\_goal}$  are the rewards for an collision with an obstacle, for the empty space in the map, and if the human is in its predicted goal location, respectively.

$$R(s) = w_{position}R_{position}(s) + w_{heading}R_{heading}(s) - w_{Pmodel}(R_{Pmodel, human}(s) + R_{Pmodel, robot}(s)) - w_{obstacle}R_{obstacle}(s) - w_{Empty}R_{Empty}(s) + w_{human\_goal}R_{human\_goal}(s) \tag{8}$$

Typically, with MMDP, each agent is assigned a goal position [43] that is given a high reward and the agents head towards their respective goals accordingly. However, since Co-MDP is a local planner that does not see the entire environment, a local goal must be selected, e.g., the highest value from the global policy in sight. This approach limits Co-MDP to only make slight deviations from the global path since the robot must reach the specified local goal, causing difficulties when the previously planned path is newly blocked. Figure 9 demonstrates this failure in Fig. 9a and c, where the robot follows the global path of Fig. 9a in c despite the path being blocked by the human. Instead, the robot could have avoided the human by going down the other hallway as shown in Fig. 9b.

**Fig. 9** The initial global plan (a), the “would be” global plan if the human was known (b), the local plan with explicit goal (c), and local plan without explicit goal (d). Blue/red denote low/high value, respectively



To increase the effectiveness of the path deviations, the local section of the provided global policy ( $\Pi_{Local}$ , which is the global optimal action direction) and global values ( $V_{local}$ ) are assigned as the foundation of Co-MDP’s reward function, where the policy direction dot multiplied by the state orientation determines the heading reward of the robot. This is shown in Eq. 10 where  $\hat{s}$  is the orientation of the individual state,  $s$  and  $\Pi_{Local}$  is the optimal action direction from the local subset of the global policy ( $\Pi_{Global}$ ), as shown in Eq. 9.

$$\Pi_{Local} \subset \Pi_{Global} \quad (9)$$

$$R_{heading}(s) = \Pi_{Local}(s) \cdot \hat{s} \quad (10)$$

The global value ( $V_{Global}$ ) determines the position reward (Eq. 12) of the robot. By doing this the robot has no explicit goal position that it must reach and therefore its “goal” becomes the highest valued state. This allows Co-MDP to make large deviations from the planned path when necessary that are influenced by global policy and the local observation. This process is demonstrated in Fig. 9.d where the robot deviated from the global path of Fig. 9.a to go up the left hallway, due to the human blocking the right hallway.

$$V_{Local} \subset V_{Global} \quad (11)$$

$$R_{position}(s) = V_{Local}(s) \quad (12)$$

This approach is only applied to the robot agent’s portion of the shared state space since only the global plan of the robot is known as the human agent is only a prediction of what the human may do. However, this does not limit Co-MDP because the human will handle its own independent long term and local path plan. Therefore, the human agent’s specific goal can be assigned as the explicit goal provided by the intention model, with a high positive reward ( $R_{Human\ goal}$ ).

The reward function is also augmented by applying large negative rewards to static objects ( $R_{Obstacle}$ ) and the preference model surrounding both the agents ( $R_{Pmodel}$ ) to prevent collisions with static objects and the agents. In addition, a small negative reward is given to empty space ( $R_{Empty}$ ) to have the agents move towards their highest value states faster.

In addition to the reward function, the transition function ( $T$ ) also heavily governs MDP behavior, because it is the probability of transitioning to the next joint state ( $s'$ ) given a joint state-action pair ( $s, a$ ). In addition, the transition function can be described as the joint probability of each agent taking its corresponding sub-action  $\langle dx, dy, d\theta \rangle$  as shown in Eq. 13. The probabilities of each agent taking each sub-action are provided by each agent’s intention specific socially-aware transition model ( $P$ ), which is displayed in Table 2.

$$T(s'|s, a) = P(\text{subAction}_{\text{robot}} \cap \text{subAction}_{\text{human}}) \quad (13)$$

Furthermore, due to the nature of the transition models existing as an integration or “build-up” approach to predicting the movements of the human, the uncertainty from the models grows exponentially with increasing time steps. Therefore, to mitigate this issue, a receding horizons approach with a three time step horizon, to match the three step intention prediction, is utilized as shown in Fig. 10. Co-MDP is re-run after every time step that another robot or human is within view. This allows the intention, predicted actions, and goals of the other agents to be updated. This reduces the long term uncertainty of the human changing intentions. Furthermore, it aids in retaining robustness for unmodeled scenarios where a human acts erratically and constantly changes intentions.

To reduce the computation time, a matrix implementation for value iteration [45] that takes advantage of linear algebra to parallelize the processes of Co-MDP is utilized.  $R$  and  $V$  are the reward and value functions expressed as vectors.  $NS$ ,

$T$ , and  $Q$  are the next state, transition, and state-action-value matrices that hold the index of the next state, transition probability, and value of the next state, respectively, when given the current state and action. The discount factor  $\gamma$  remains a scalar value. These vectors and matrices are connected using the Bellman Equation (Eq. 14) as shown in Fig. 8

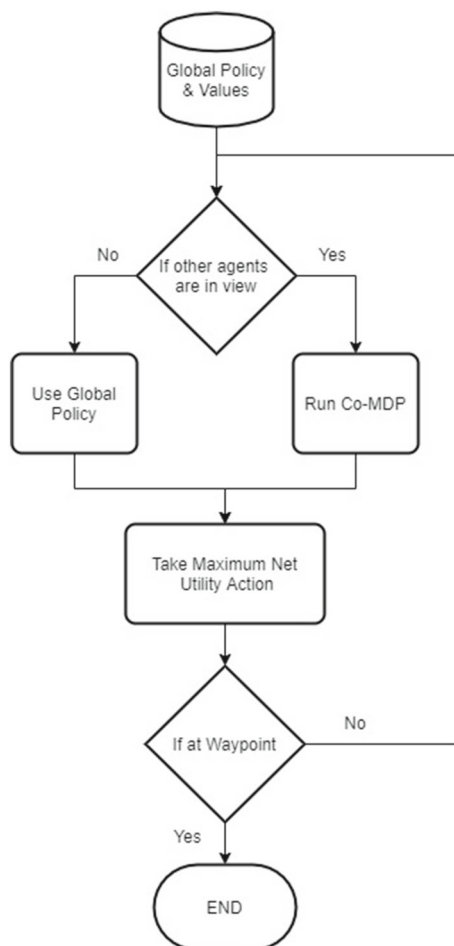
$$V(s) = \max(R(s, a) + \gamma \sum P(s, a, s')V(s')) \quad (14)$$

### 3.4 Experiment Design with Human -in-the-Loop Assessment

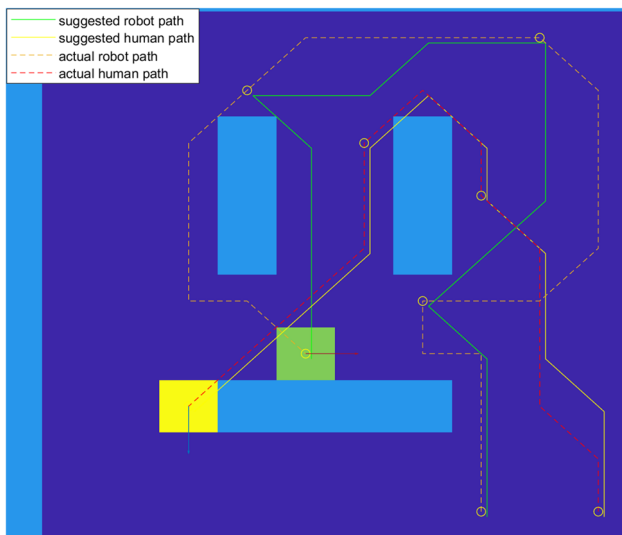
Due to the subjective nature of social interaction rules, it is difficult to extend the verification of following these rules past direct observation. Therefore, to verify the performance and the anthropomorphic obstacle avoidance behavior of Co-MDP, a human-in-the-loop approach was designed to test if the trajectories generated from the proposed algorithm or from the human operators can be distinguished by human observers. The proposed algorithm is evaluated by the percentage of human observers that identify the algorithm as a human and the confidence of their answer. A simulated supermarket scenario was used to demonstrate the obstacle avoidance behavior of two agents (human or robot). There were two phases of the experiment: (1) data generation, and (2) socially-aware obstacle avoidance behavior verification.

*Phase I, Data Generation*—for this portion of the experiment, four experienced robotics developers with no prior knowledge of the proposed algorithm were recruited as the human operators to generate the simulated human trajectories, their mean $\pm$ std age was  $28.0 \pm 2.2$ . By selecting roboticists for this task, it was hoped that participants would be able to emulate behaviors as close to their innate navigational behaviors as possible. Lay participants would be presented with multiple distractions not present in a motion capture setting: understanding an occupancy grid, perceiving an abstract agent as one’s “self” and using a keyboard to navigate accurately. As a result, their navigation in this test may not be representative of their behavior in a real-world environment. Roboticists are more likely to be comfortable with the side-effects of the testing environment, and thus may be better able to perform naturally on the navigation task. Before the experiment, adequate training and time were given to the operators before they felt comfortable to generate data. Once the experiment commenced, the instructions provided to the operators was,

*You are to behave as if you are walking and picking four items in a supermarket. While staying as close to the pre-defined path for each segment as possible without colliding with the other agent, which could be either a human or a robot.*



**Fig. 10** Receding horizons in waypoint navigation example with Co-MDP



**Fig. 11** The final image from one of the experiment videos with the suggested and actual paths taken by both agents

The predefined path was the shortest path towards the next item and served as a reference so that the operators could focus on the obstacle avoidance rather than the path planning. Each operator was asked to accomplish five trials with four goals in each. During each trial, the trajectory generated by the operator, the trajectory of the other agent, and the predefined shortest paths of the two agents were recorded. An example of this experiment is shown in Fig. 11, which displays all the suggested predefined paths to all the goals and actual paths taken by the agents.

*Phase II, Socially-Aware Obstacle Avoidance Behavior Verification*—21 participants, 10 males and 11 females, were recruited to conduct the second phase of the experiment, their mean  $\pm$  std age was  $33.3 \pm 12.9$ . Seventeen participants reported to have no experience with robot obstacle avoidance algorithms and four participants reported to have 1–2 years of experience. In each experimental trial, participants were asked to watch a video showing the obstacle avoidance behavior of two agents interacting in the same scene and then identify which agent(s) were controlled by humans. Each participant watched eight videos (2 modes of agent A (human or Co-MDP operated)  $\times$  2 modes of agent B  $\times$  2 repetitions). In addition, participants were asked about their confidence in the predictions after each trial from 1 to 5, and the confidence level was normalized to 20–100%. The six exact questions are:

- In this video, do you think the YELLOW block was a Robot or Human?  
A: yes / no
- Please rate your confidence in your answer to the previous question.  
A: (I'm not sure) 1–5 (I'm very confident)

- In this video, do you think the GREEN block was a Robot or Human?  
A: yes / no
- Please rate your confidence in your answer to the previous question.  
A: (I'm not sure) 1–5 (I'm very confident)
- How well did the blocks avoid each other?  
A: (poorly (collided together, took significantly more steps than the path, etc.)) 1–5 (excellent (effortlessly, human-like, etc.))
- How predictable were the movements of the blocks?  
A: (not predictable at all) 1–5 (fully predictable)

By the end of the experiment, they were also given the opportunity to share what agent motion features they leveraged for their decision making process. To further diversify the testing dataset for better generalizability of the results, we prepared four different sets and randomly assigned one to each participant.

After data collection, the participant's responses were compared to the truth. Furthermore, the confidence of the participants for each prediction was also compared and analyzed. Statistical analysis was conducted with statistical significance achieved when  $p < 0.05$ . Due to the effects of COVID-19, all experiments were conducted online without face-to-face contact. The experiment protocol was approved by the University of Florida Institutional Review Board (IRB202003142).

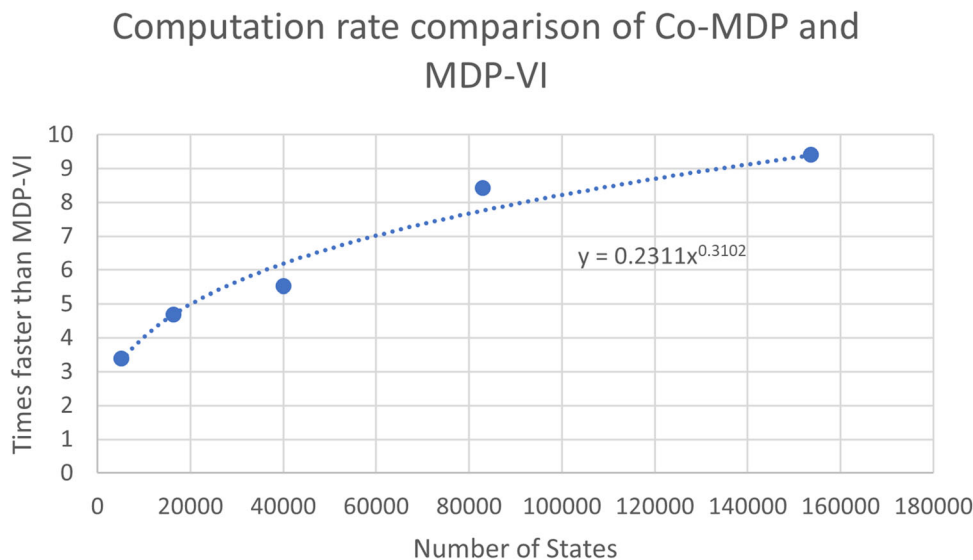
## 4 Results and Discussion

For the proposed algorithm, using the human motion models with Co-MDP, the simulated robots were observed to follow the provided social norms, and exhibit socially-aware behavior. These behaviors were verified through simulation testing and human participant evaluation of the anthropomorphic behavior.

The first simulation testing was computation time. Implementing Co-MDP as a matrix had the effect of reducing the rate of computation by approximately nine times the rate of standard value iteration (VI) for our used number of states 153,664, as shown in Fig. 12.

The rate of collision for Co-MDP was assessed using the human-in-the-loop data. Throughout each of the four goals of the twenty videos (eighty trials) used in the human participant experiment, only one collision, where both agents occupied the same grid cell, occurred between a human operator and the robot. This occurred when the robot was against a wall and the human operator violated expected personal space norms, coming very close to the robot. Thus the collision rate for this algorithm is 1.25% and it shows the feasibility of using

**Fig. 12** Rate of computation comparing matrix implemented Co-MDP with standard value iteration (MDP-VI). Where the solid dots are the raw experimental results, and the dotted line is an exponential fit line



Co-MDP as an obstacle avoidance algorithm in environments with a human, with room for future improvements.

Another simulation test that validates a multitude of social interactions such as passing on the left, providing adequate personal space, and cooperation is the traditional hallway test [7]. The hallway test (shown in Fig. 13) consists of two agents one controlled by Co-MDP (green) and the other controlled by a human (yellow), that are to reach a predefined destination at the other end of the hallway by passing each other on the left (for right hand traffic countries) with adequate personal space. Co-MDP repeatedly expresses both of these behaviors.

In addition to the hallway scenario, the supermarket scenario used by the human participants also exhibits socially-aware behavior. For example, as shown by the paths in Fig. 11, the robot deviated from its suggested shortest path to ensure adequate personal space around the human throughout the entire experiment, except when prevented by environmental factors such as tight hallways. In addition, the robustness of the algorithm to find a new path to maximize social utility is shown by the robot traversing around the aisle to provide adequate personal space despite the center hallway being a shorter path.

From the human participant verification of the algorithm, a confusion matrix demonstrating how participants classified agents in different cases, along with the confidence of the participants' answers are shown in Table 3 and Fig. 14. Identification results of 21 participants are shown in Table 3. As shown in the confusion matrix, among 168 observations that the agent was robot, 90 of them (53.57%) were labeled as human agents by participants; while among 168 observations that the agent was human, only 69 (41.07%) were labeled as humans. One-tailed two-proportion z-test revealed that the percentage of labeling a true robot agent as a human (53.57%) is significantly higher than labeling a true human

**Table 3** Confusion matrix of participants' identification results

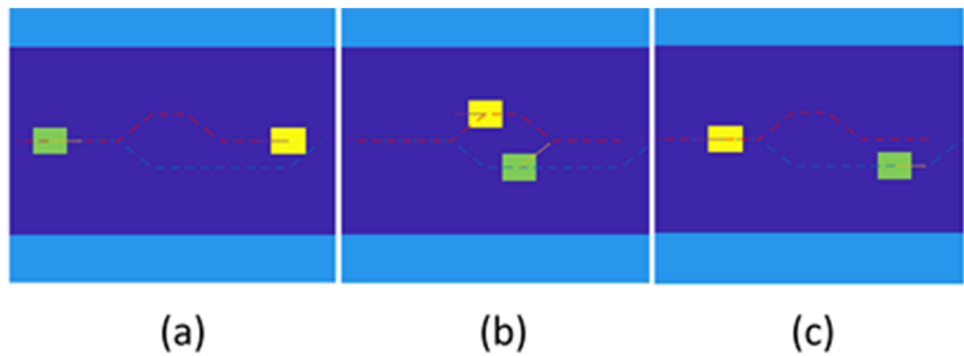
		Prediction		
		Robot	Human	
Truth	Robot	78	90	168
	Human	99	69	168
		177	159	

Truth denotes which agent type was observed, robot or human, and Prediction denotes which type the participant identified it as

agent as a human (41.07%), with the z-value = 2.29 and  $p = 0.01 < 0.05$ .

Collectively, these results indicated that the socially-aware Co-MDP we proposed achieved the expected performance in the simulated testing environment, and showed substantial human-like behavior. However, these results should not be overstated. At face value, the Co-MDP agents even outperformed human agents in the test (21 more observations were labeled as humans). There are many factors that may have contributed to this outcome: (1) although operators we recruited were all seasoned robotics developers with at least 3 years experience, and adequate training sessions were provided for them to learn how to operate the agent, it is possible that their operation did not fully represent their intention and their performance deteriorated, and (2) the testing environment was a simplification of the real-world scenario, which made the environment perception more challenging on the human side than the algorithm side. In terms of the confidence of identification, as shown in Fig. 14, the average confidence of the correct identification (73.5%) is higher than the false identification (69.0%). More specifically, the average confidence of identifying a robot agent correctly (75.7%) was higher than a false recognition (67.8%); similarly, the average confidence of successfully identifying a

**Fig. 13** Human–robot interaction example in a hallway environment. The human (yellow) and robot (green) pass in opposite directions, with frames taken at the beginning (a) middle (b) and end (c) of the task. Light blue denotes static obstacles, dark blue denotes free space



human agent (70.5%) was higher than a false recognition (69.7%). The higher level of confidence of correct identification indicated that the predictions from participants were backed by different evidence rather than random guesses. Overall, the results from this human-in-the-loop experiment were promising, which provide further support for the feasibility of our proposed approach. Follow up assessments with physical human–robot interaction will be necessary to evaluate the algorithm at a higher fidelity level.

From the descriptions of the criteria the participants used to identify the agents as humans or robots an unprompted dichotomous viewpoint emerged. A majority of the participants (14 of 21) had the viewpoint that humans would be more passive by taking longer paths to avoid the robot, while the robot would follow its path more directly. Six other participants believed the opposite to be true. This is further supported by Fig. 14 where the average confidence of the participants for identifying the robot as either human or a robot was high at approximately 70%. This finding illustrates that a universal socially-aware robot behavior algorithm is challenging to develop and this direction of research merits further investigation.

Another interesting observation is that the participants identified the human agents as robots more often than as humans. This result may be due to the human operators potentially viewing the scenario as a game, where they have to minimize their path as much as possible and since it is in simulation, they can be more direct or “robotic-like” according to our human participants than if they were in the real-world with the robot reacting naturally without explicitly thinking about their actions. In addition, since the selected human operators were experienced robotics developers they may be biased towards behaving more robotic-like (i.e., taking action to maximize utility).

During the experiment, participants were asked to rate the collective obstacle avoidance performance after each video on a 5-point Likert-scale from 1-poorly to 5-excellent. Results of participants’ subjective rating are shown in the

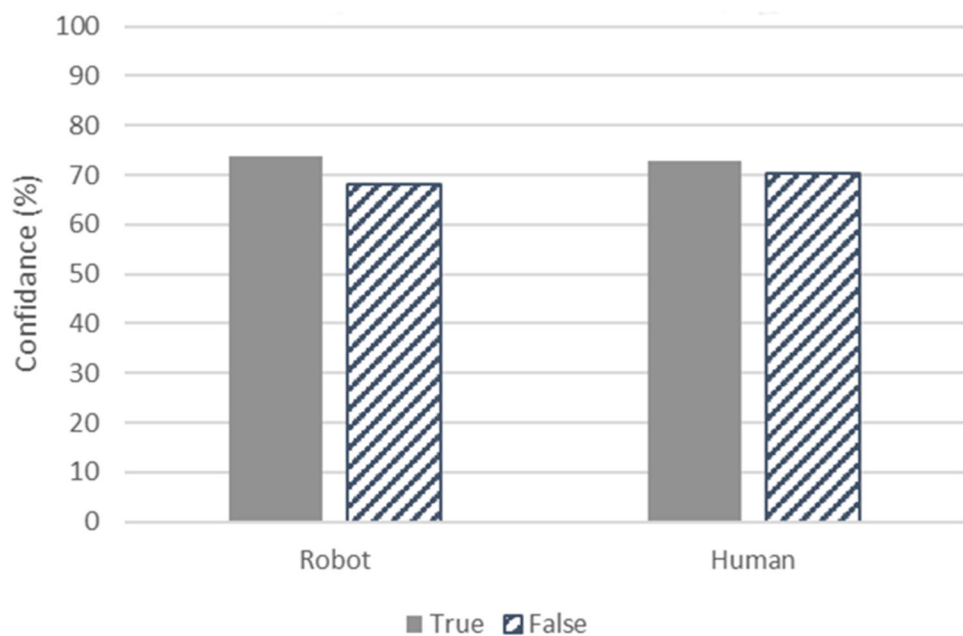
**Table 4** Participants’ ratings of the subjective performance of collective obstacle avoidance, grouped by the type of interaction observed

Interaction	Mean (std) performance
Robot–robot (RR)	3.83 (0.82)
Human–human (HH)	3.67 (0.79)
Human–robot (HR)	3.29 (1.16)

Table 4, where the condition refers to the ground truth labels of the agents in the videos. Results from Kruskal–Wallis H Test revealed that there was a significant difference between the participants’ rating of the three conditions from (H-value = 7.083,  $p = 0.029$ ). Mann–Whitney U Test revealed that the average rating of RR was significantly different from the average rating of HR ( $p = 0.015$ ). No significant difference was observed between RR–HH ( $p = 0.413$ ), and HR–HH ( $p = 0.088$ ).

In the simulation study the case where both agents were robots can be considered an socially aware MMDP comparison of the agents. This is because both robots predict not only the intention of the other robots, but also what the other robot would think its intention is. The reason for this is in order to act humanly, the robots must move in ways that others expect it to move, so each robot uses the transition model that the other agent thinks it has. Therefore, since both agents are robots and they use the intention model that the other agent thinks it has, both agent’s predictions of the other’s intention are true. Thus each robot knows the true intention of the other, and they would behave the same as if they were in a direct MMDP algorithm. However there is a deviation from standard MMDP, since both robot’s local worlds are egocentric, their shared state-spaces are slightly different. Based on Table 4 it is shown that the two robots interaction performed only slightly higher than the other cases. The higher performance may be attributed to this advantage that each robotic agent shares the same models and can perfectly predict the true intention of the other agent, which is not possible in the human interaction cases.

**Fig. 14** Participants' confidences in their identifications



## 5 Conclusion

This paper presents an interdisciplinary effort combining human studies, robotics, and machine learning knowledge in developing a socially-aware obstacle avoidance algorithm.

The first contribution of this work is the consolidation and development of discretized probabilistic short-term human motion models that can be easily implemented into robotic algorithms. The result of these models allows the improvement of predictive socially acceptable cooperative obstacle avoidance between groups of humans and robots in the form of Co-MDP. In addition, a verification procedure of the anthropomorphic behavior of Co-MDP was developed, using a novel experimental design with human-in-the-loop simulations and human participant observation verification through statistical study.

## 6 Limitations and Future Work

The main limitation of Co-MDP, like MMDP, is the exponential growth in state space with an increase in the number of agents or size of the local map. For instance, using Eq. 6 with two agents results in 153,664 states. However, with three agents, it scales to over 60 million states. Additionally, the verification process of this algorithm was limited to simulations in a simplified grid world with only two active agents. Another limitation was that the numeric values in the human intention model and transition model were empirically selected in this study to represent a proof of concept. Human studies towards a better understanding of the uncertainty of human intention and motion variance are needed to make

the human models more representative of rational human behaviors.

Future work would include alleviating the computational limitations of the Co-MDP, while also maintaining the target computation rate of 1 Hz. One alleviating method would be approximating the very large Q-matrix with the deep reinforcement learning techniques as demonstrated by [22–25]. This could be further improved by pruning the possible actions that the agent cannot take based on the Human Transition Models, from the action space. This would reduce the size of the T and Q matrices by 1/3, if the intention is not "Random" which would decrease the computation time [50]. Another future direction is to expand this framework to more agents, based on the principle of fast decaying interaction effects with distance [46], by only considering  $k$  neighbors.

To further improve the simulated testing motion capture of humans and robots interacting will be done to generate real-world data that captures the implicit actions of humans. In addition, a physical experiment will be conducted by having a mobile robot navigate in a room with humans and observing not only the robot's behavior but the humans' behavior as well, to verify the social-awareness of Co-MDP at a higher fidelity level.

This testing would also provide better understanding of human intention and motion variance, by having the robot learn the human transition and intention models using a model free Q-learning approach similar to learned approaches in [51]. Thus, improving the fidelity of the motion models. Furthermore, an analytical study of the reward weighing parameters similar to [52] would increase

the fidelity of the empirically Co-MDP reward derived weights.

In addition, the intention selection method could be improved with a Bayesian approach as [53] uses to predict what object a human is intending to pick up. The reason why a Bayesian approach was not initially included was due to the absence of a model of how human intentions change over time. Thus investigations on the change of human intentions over time will be pursued in future studies.

The outcomes of the human-in-the-loop experiment also suggest additional research directions in understanding how people perceive a robot's behavior as socially acceptable and how improvements in robot's decision making capabilities affect human behaviors around these robots. In the current implementation of the human models, more specifically, the cost functions used in the human preference models, the only variable being considered to represent the cost changes caused by human motions was the relative moving velocity between the human and the robot  $v$ , and its value was set to be constant for simplification. In addition to the velocity the emotional state of the human would change the cost distribution surrounding it, due to the change in behavior. Studies have been conducted by [54] on how robots can convey and perceive social queues to determine the emotional state of the human, and how their behavior changes. As for the next step, the real-time changes of the preference model will be included, based on the current relative velocity and the emotional state of the human. In addition, it would be interesting to investigate if the robot can benefit from assigning different cost functions (e.g. shape, size, and cost values) around the human based on his/her intentions and motions and generate more socially aware obstacle avoidance behaviors. Furthermore, [55] provides a compilation of eight areas of social cognition some of which were addressed in this paper, such as "leverage simulation-based top-down perceptual biasing", which allow for extrapolation of the current state into predicted future states (Algorithm 1). This is in addition to other areas of social cognition that were not included such as "provide motor and perceptual resonance mechanisms" that allow for continued learning of HRI, to be included into the future work.

## Declarations

**Funding** Partial support was provided by West Virginia University Statler College Undergraduate Research Funds, US National Science Foundation Research Experience for Undergraduates Program, Award #1851815, and the University of Florida Faculty Startup Funds.

**Conflict of interests** The authors declare that they have no conflict of interest.

**Ethical standards** The human-in-the-loop study performed in this work was approved by the University of Florida Institutional Review Board, IRB202003142.

**Financial interests** The authors declare they have no financial interests related to this work.

## References

1. Lasota PA, Shah JA (2015) Analyzing the effects of human-aware motion planning on close-proximity human-robot collaboration. *Hum Factors* 57(1):21–33
2. Paulin R, Fraichard T, Reignier P (2019) Using human attention to address human–robot motion. *IEEE Robot Autom Lett* 4(2):2038–2045
3. Chandan G, Jain A, Jain H, Mohana (2018) Real time object detection and tracking using deep learning and opencv. In: 2018 International Conference on inventive research in computing applications (ICIRCA), pp 1305–1308. <https://doi.org/10.1109/ICIRCA.2018.8597266>
4. Brousseau B, Rose J (2012) An energy-efficient, fast fpga hardware architecture for opencv-compatible object detection. In: 2012 International conference on field-programmable technology. IEEE, pp 166–173
5. Ulrich I, Borenstein J (2000) Vfh/sup \*/: local obstacle avoidance with look-ahead verification. In: Proceedings 2000 ICRA. Millennium conference. IEEE international conference on robotics and automation. Symposia proceedings (Cat. No.00CH37065), vol 3. pp 2505–2511, <https://doi.org/10.1109/ROBOT.2000.846405>
6. Hart PE, Nilsson NJ, Raphael B (1968) A formal basis for the heuristic determination of minimum cost paths. *IEEE Trans Syst Sci Cybern* 4(2):100–107. <https://doi.org/10.1109/TSSC.1968.300136>
7. Pacchierotti E, Christensen HI, Jensfelt P (2005) Human–robot embodied interaction in hallway settings: a pilot user study. In: ROMAN 2005. IEEE international workshop on robot and human interactive communication, 2005. IEEE, pp 164–171
8. Minguez J, Montano L (2004) Nearness diagram (nd) navigation: collision avoidance in troublesome scenarios. *IEEE Trans Robot Autom* 20(1):45–59
9. Ogren P, Leonard NE (2005) A convergent dynamic window approach to obstacle avoidance. *IEEE Trans Robot* 21(2):188–195. <https://doi.org/10.1109/TRO.2004.838008>
10. Fox D, Burgard W, Thrun S (1997) The dynamic window approach to collision avoidance. *IEEE Robot Autom Mag* 4(1):23–33. <https://doi.org/10.1109/100.580977>
11. Warren, CW (1989) Global path planning using artificial potential fields. In: 1989 IEEE international conference on robotics and automation. IEEE Computer Society, pp 316–317
12. Warren, CW (1990) Multiple robot path coordination using artificial potential fields. In: Proceedings, IEEE international conference on robotics and automation, vol 1. pp 500–505. <https://doi.org/10.1109/ROBOT.1990.126028>
13. Vadakkepat P, Tan KC, Wang M-L (2000) Evolutionary artificial potential fields and their application in real time robot path planning. In: Proceedings of the 2000 congress on evolutionary computation. CEC00 (Cat. No.00TH8512), vol 1. pp 256–263. <https://doi.org/10.1109/CEC.2000.870304>
14. Sisbot EA, Clodic A, Fontmarty M, Brethes L, Alami, R et al (2006) Implementing a human-aware robot system. In: ROMAN 2006—The 15th IEEE international symposium on robot and human interactive communication. IEEE, pp 727–732
15. Chung SY, Huang HP (2011) Predictive navigation by understanding human motion patterns. *Int J Adv Robot Syst* 8(1):3



16. Ferrer G, Sanfeliu A (2014) Proactive kinodynamic planning using the extended social force model and human motion prediction in urban environments. In: 2014 IEEE/RSJ international conference on intelligent robots and systems. IEEE, pp 1730–1735
17. Chung SY, Huang HP (2010) A mobile robot that understands pedestrian spatial behaviors. In: 2010 IEEE/RSJ international conference on intelligent robots and systems. IEEE, pp 5861–5866
18. Bai H, Cai S, Ye N, Hsu D, Lee WS (2015) Intention-aware online pomdp planning for autonomous driving in a crowd. In: 2015 IEEE international conference on robotics and automation (icra). IEEE, pp 454–460
19. Foka A, Trahanias P (2010) Probabilistic autonomous robot navigation in dynamic environments with human motion prediction. *Int J Soc Robot* 2:79–94. <https://doi.org/10.1007/s12369-009-0037-z>
20. Madani, O, Hanks, S, Condon, A (1999) On the undecidability of probabilistic planning and infinite-horizon partially observable markov decision problems. In: AAI/IAAI. pp 541–548
21. Tsitskilis J, Papadimitriou C (1987) The complexity of Markov decision processes. *Math Oper Res* 12(3):441–450
22. Chen YF, Everett M, Liu M, How JP (2017) Socially aware motion planning with deep reinforcement learning. In: 2017 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp 1343–1350
23. Mnih V, Kavukcuoglu K, Silver, D, Graves, A, Antonoglou, I, Wierstra, D, Riedmiller, M (2013) Playing atari with deep reinforcement learning. [arXiv:1312.5602](https://arxiv.org/abs/1312.5602)
24. Hessel, M, Modayil, J, Van Hasselt, H, Schaul, T, Ostrovski, G, Dabney, W, Horgan, D, Piot, B, Azar, M, Silver, D (2017) Rainbow: combining improvements in deep reinforcement learning. [arXiv:1710.02298](https://arxiv.org/abs/1710.02298)
25. Karkus, P, Hsu, D, Lee, WS (2017) Qmdp-net: Deep learning for planning under partial observability. In: Advances in neural information processing systems. pp 4694–4704
26. Cai, P, Luo, Y, Saxena, A, Hsu, D, Lee, WS (2019) Lets-drive: Driving in a crowd by learning from tree search. [arXiv:1905.12197](https://arxiv.org/abs/1905.12197)
27. Chandra S, Bharti AK (2013) Speed distribution curves for pedestrians during walking and crossing. *Procedia Soc Behav Sci* 104:660–667
28. Taylor MJD, Dabnichki P, Strike S (2005) A three-dimensional biomechanical comparison between turning strategies during the stance phase of walking. *Hum Mov Sci* 24(4):558–573
29. Gupta, A, Johnson, J, Fei-Fei, L, Savarese, S, Alahi, A (2018) Social gan: socially acceptable trajectories with generative adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 2255–2264
30. Schulz AT, Stiefelhagen R (2015) A controlled interactive multiple model filter for combined pedestrian intention recognition and path prediction. In: 2015 IEEE 18th international conference on intelligent transportation systems. IEEE, pp 173–178
31. Rabinowitz, N, Perbet, F, Song, F, Zhang, C, Eslami, SMA, Botvinick, M (2018) Machine theory of mind. In: Dy J, Krause A (eds) Proceedings of the 35th international conference on machine learning, proceedings of machine learning research, vol 80. PMLR, pp 4218–4227
32. Pynadath, DV, Marsella, SC (2005) Psychsim: Modeling theory of mind with decision-theoretic agents. In: Proceedings of the 19th international joint conference on artificial intelligence, IJCAI'05. Morgan Kaufmann Publishers Inc., San Francisco, pp 1181–1186
33. Petković T, Marković I, Petrović I (2017) Human intention recognition in flexible robotized warehouses based on markov decision processes. In: Iberian robotics conference. Springer, pp 629–640
34. Petković T, Puljiz D, Marković I, Hein B (2019) Human intention estimation based on hidden markov model motion validation for safe flexible robotized warehouses. *Robot Comput Integr Manuf* 57:182–196
35. Baker CL, Tenenbaum JB (2014) Modeling human plan recognition using bayesian theory of mind. In: Plan, activity, and intent recognition: theory and practice. pp 177–204
36. Sisbot EA, Marin-Urias LF, Alami R, Simeon T (2007) A human aware mobile robot motion planner. *IEEE Trans Robot* 23(5):874–883
37. Rios-Martinez J, Renzaglia A, Spalanzani A, Martinelli A, Laugier C (2012) Navigating between people: a stochastic optimization approach. In: 2012 IEEE international conference on robotics and automation. IEEE, pp 2880–2885
38. Scandolo L, Fraichard T (2011) An anthropomorphic navigation scheme for dynamic scenarios. In: 2011 IEEE international conference on robotics and automation. IEEE, pp 809–814
39. Hall ET (1966) The hidden dimension, vol 609. Doubleday, Garden City
40. Kirby R, Simmons R, Forlizzi J (2009) Companion: a constraint-optimizing method for person-acceptable navigation. In: RO-MAN 2009—the 18th IEEE international symposium on robot and human interactive communication. IEEE, pp 607–612
41. Boutilier C (1999) Sequential optimality and coordination in multiagent systems. *IJCAI* 99:478–485
42. Proper S, Tadepalli P (2009) Solving multiagent assignment markov decision processes. In: Proceedings of the 8th international conference on autonomous agents and multiagent systems, vol 1. pp 681–688
43. Lauer M, Riedmiller M (2000) An algorithm for distributed reinforcement learning in cooperative multi-agent systems. In: Proceedings of the seventeenth international conference on machine learning. Morgan Kaufmann, pp 535–542
44. Becker R, Zilberstein S, Lesser V, Goldman CV (2004) Solving transition independent decentralized Markov decision processes. *J Artif Intell Res* 22:423–455
45. Sapio A, Bhattacharyya SS, Wolf M (2018) Efficient solving of Markov decision processes on gpus using parallelized sparse matrices. In: 2018 conference on design and architectures for signal and image processing (DASIP). IEEE, pp 13–18
46. Qu G, Li N (2019) Exploiting fast decaying and locality in multiagent mdp with tree dependence structure. In: 2019 IEEE 58th conference on decision and control (CDC). pp 6479–6486. <https://doi.org/10.1109/CDC40024.2019.9029635>
47. Pentland A, Liu A (1999) Modeling and prediction of human behavior. *Neural Comput* 11(1):229–242. <https://doi.org/10.1162/089976699300016890>
48. Gérin-Lajoie M, Richards CL, McFadyen BJ (2005) The negotiation of stationary and moving obstructions during walking: anticipatory locomotor adaptations and preservation of personal space. *Mot Control* 9(3):242–269
49. Kruse T, Pandey AK, Alami R, Kirsch A (2013) Human-aware robot navigation: a survey. *Robot Auton Syst* 61(12):1726–1743
50. Pinto J, Fern A (2017) Learning partial policies to speedup mdp tree search via reduction to iid learning. *J Mach Learn Res* 18(65):1–35
51. Tutsoy O, Barkana DE (2021) Model free adaptive control of the under-actuated robot manipulator with the chaotic dynamics. *ISA transactions*
52. Tutsoy O, Brown M (2016) An analysis of value function learning with piecewise linear control. *J Exp Theor Artif Intell* 28(3):529–545
53. Jain S, Argall B (2019) Probabilistic human intent recognition for shared autonomy in assistive robotics. *ACM Trans Hum Rob Interact* THRI 9(1):1–23
54. Fiore SM, Wiltshire TJ, Lobato EJ, Jentsch FG, Huang WH, Axelrod B (2013) Toward understanding social cues and signals in human-robot interaction: effects of robot gaze and proxemic behavior. *Front Psychol* 4:859

55. Wiltshire TJ, Warta SF, Barber D, Fiore SM (2017) Enabling robotic social intelligence by engineering human social-cognitive mechanisms. *Cogn Syst Res* 43:190–207

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.