# MIRROR: Differentiable Deep Social Projection for Assistive Human-Robot Communication


***Abstract*—Communication is a hallmark of intelligence. In this work, we present MIRROR, an approach to (i) quickly learn human models from human demonstrations, and (ii) use the models for subsequent communication planning in assistive shared-control settings. MIRROR is inspired by social projection theory, which hypothesizes that humans use self-models to understand others. Likewise, MIRROR leverages self-models learned using reinforcement learning to bootstrap human modeling. We discuss a human-subject study using the CARLA simulator which shows that (i) MIRROR is able to scale to complex domains with high-dimensional observations and complicated world physics and (ii) provides effective assistive communication that enabled participants to drive more safely in adverse weather conditions.**


## I. Introduction

Communication is an essential skill for intelligent agents; it facilitates cooperation and coordination, and enables teamwork and joint problem-solving. However, effective communication is challenging for robots; given a multitude of information that can be relayed in different ways, how should the robot decide *what*, *when*, and *how* to communicate?

In this paper, we focus on assistive shared-control or teleoperation scenarios. As an example, consider the scenario in Fig. 1 where a robot assistant is tasked to provide helpful information to a human driving the blue car in fog (or to explain the robot's own driving behavior). There are other cars in the scene, which may not be visible to the human. The robot, however, has access to sensor readings that reveal the environment and surrounding cars. To prevent potential collisions, the robot needs to communicate relevant information — either using a heads-up display or verbally — that intuitively, should take into account what the human driver currently believes, what they can perceive, and the actions they may take.

Prior work on planning communication methods in HRI typically rely on human models, which are typically handcrafted using prior knowledge (e.g., [1], [2]) or learned from collected human demonstrations (e.g., [3]). Unfortunately, handcrafted models do not easily scale to complex real-world environments with high-dimensional observations, and data-driven models typically require a large number of demonstrations to generalize well. In this work, we seek to combine prior knowledge with data in a manner that reduces *both* manual specification and sample complexity.

Our key insight is that learning *differences* from a suitable reference model is more data-efficient than learning an entire human model from scratch. We take inspiration from *social projection theory* [4], which suggests that humans have a tendency to expect others to be similar to ourselves, i.e., a person understands other individuals using one's self as a

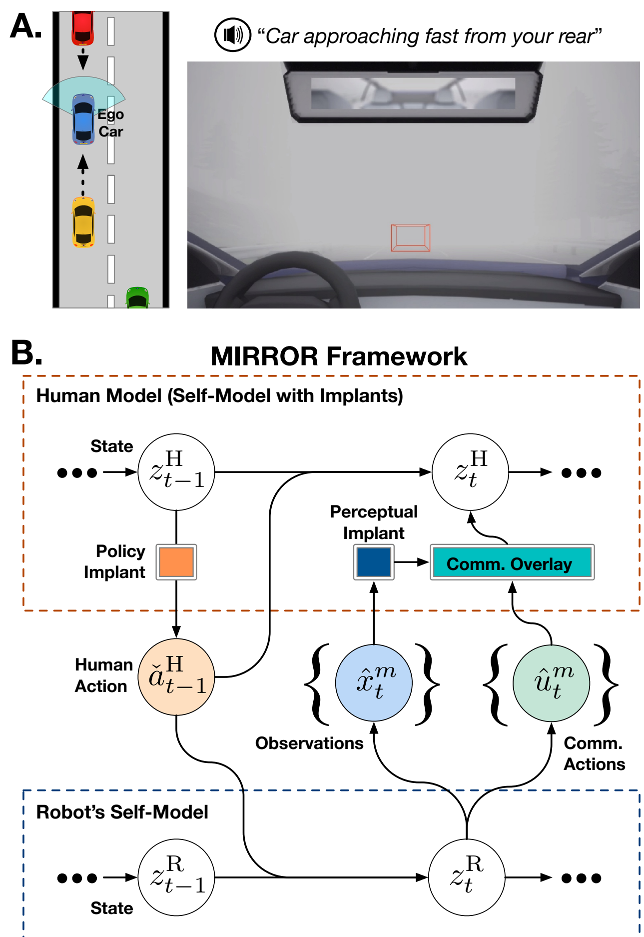


Fig. 1: Human-Robot Communication Example. (**A**) A Robot Assistant needs to provide information to help a human driving the blue car in dense fog. The human has limited visibility and the assistant can highlight cars on a heads-up-display or provide verbal cues (as the human is unable to see highlighted cars that are in the rear). In this scenario, a collision is imminent—the red car in front is slowing down and a yellow car speeding up from the rear. Our proposed MIRROR assistant immediately highlights the red car in front and verbally tells the user about yellow car at the rear. It chooses *not* to tell the human about the green car, whose location has little impact on the human's decision. (**B**) Our MIRROR framework is inspired by *social projection*; the robot reasons using a human model that is constructed from its own internal self-model. Strategically-placed learnable "implants" capture how the human is different from the robot. MIRROR plans communicative actions by forward simulating possible futures by coupling its own internal model (to simulate the environment) and the human model (to simulate their actions).

reference. This inductive bias can be effective when the agents are similar and may promote cooperation [5], [6]. From a cognitive perspective, social projection is a heuristic by which we evaluate and predict another agent's behavior. Likewise, a robot can use its self-model to reason about a human; in our setup, the robot first learns how to perform the task on its own then uses this model to reason about other agents.

A natural concern is that the human is unlikely to be similar to the robot; indeed, recent work has shown that (model-free) RL agents that are trained to respond to self-policies do not work well with actual humans [7]. The ways in which the human and robot perceive the world and make decisions are likely to differ. Here, we aim to isolate and learn these differences in a sample-efficient manner.

We exploit latent state-space models obtained via deep reinforcement learning and strategically place *learnable implants* that capture differences in perception and/or policy (Fig. 1.B.). These implants can be small and relatively unstructured (e.g., neural networks) or specified using prior knowledge about the human (e.g., known cognitive biases or perceptual limitations) with associated parameters that can be quickly optimized via gradient-based learning.

We call our framework **M**odel **I**mplants for **R**apid **R**eflective **O**ther-agent **R**easoning/Learning (MIRROR); an allusion to the mirror neurons in human brains that are hypothesized to play a role in social projection [8], [9]. Returning to our example, our MIRROR-enabled robot highlights the red car in front and verbally informs the user about yellow car at the rear. To avoid distracting the user, it chooses *not* to tell the human about the green car, whose location and velocity has little impact on the human's decision and safety. These communicative actions are the result of *planning* over the implant-augmented self-model and the robot's internal model (Fig. 1.B.).

We conducted a human-subject study ($n = 21$) using the CARLA simulator [10], which reveals that MIRROR provides useful assistive information, enabling participants to complete a driving task with fewer collisions in adverse visibility conditions. In summary, this abstract makes the following contributions:

- MIRROR, a sample-efficient framework for learning human models using deep self-models for initial structure;
- A planning-based communication approach that leverages learned world dynamics and human models;
- Findings from a human-subject study in the assistive driving domain, showing that MIRROR provides useful communication that improves task performance.

We believe MIRROR is a step towards better data-efficient human models for human-robot interaction; to our knowledge, MIRROR is the first work to demonstrate how deep representation learning during RL can be combined with demonstrations for human modeling and planning. MIRROR can be used for human-robot communication during robot tele-operation and shared-control settings, and it opens up an alternative path to human-robot collaboration with deep models.

## II. BACKGROUND & RELATED WORK

MIRROR builds upon the existing literature on human modeling and human-robot communication. Due to its importance, the field of agent communication is large; here, we briefly summarize closely-related work that learn and use human models for human-robot/AI interaction and communication [7], [11]–[13].

**Human Models for HRI.** In this work, we focus on model-based methods that explicitly model human behavior. Compared to model-free approaches to HRI, model-based methods tend to make reasonable predictions with far less data [12]. Model-based methods can be "black-box" in that they make few assumptions about the human and focus on learning a policy function. For example, recent work learns a human policy via imitation learning, followed by a residual policy for shared control [14]. In contrast, Theory of Mind (ToM) models incorporate (possibly strong) assumptions about how humans perceive the world and make decisions. For example, a ToM model may assume people are rational and learn in a Bayesian manner [15]–[17], which is generally not true [16].

MIRROR can be seen as a hybrid approach that scaffolds human model learning using the robot's own internal model (obtained using RL). Compared to standard black-box human models, MIRROR provides additional structure that can ease data requirements. Compared to handcrafted ToM approaches [2], MIRROR is able to handle high-dimensional observations. MIRROR is related to recent approaches that focus on capturing human traits, e.g., biases under risk and uncertainty [18] or action errors due to misunderstood environmental dynamics [19]. However, these approaches typically build on top of hand-crafted ToM models.

**Assistance via Human-Robot Communication.** Enabling robots to communicate with humans has had a long history; early robots in the 1990s (e.g., Polly [20] and RHINO [21]) were simple stimulus-response systems. In contrast, modern day robots leverage learning and planning to generate a variety of communication patterns, e.g., legible motion [22]–[24], and natural language [25], [26].

Recent work has shown that human-robot communication can improve human task performance [1], [27], explain robot errors [28] and calibrate human-robot trust [29]. However, these approaches typically use hand-specified human models and known environment models. In contrast, MIRROR plans communication actions using *learned* models. MIRROR is related to prior work on personalized assistive navigation [30], but the mechanism differs: MIRROR adapts implant parameters whilst [30] uses a mixture of expert models.

MIRROR is closely-related to Assistive State Estimation (ASE) [3] in that both approaches augment user observations to communicate state information. However, there are crucial differences: ASE assumes known dynamics and perceptual models to compute the near-optimal human policies, while MIRROR uses learned dynamics and implant models. In addition to differences in the human model, ASE aligns the user's belief and the assistant's belief. Instead, MIRROR forward

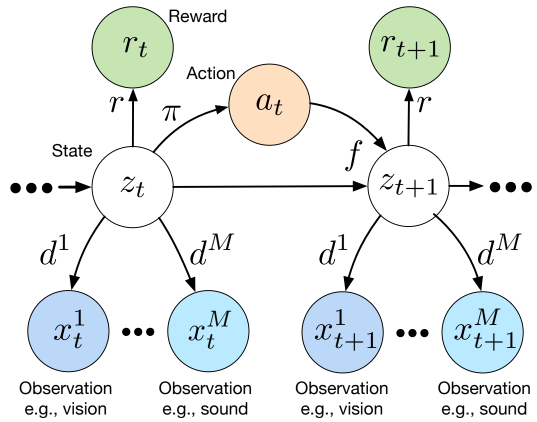


Fig. 2: MIRROR's self-model is a multi-modal latent state-space model (MSSM). In the above, circle nodes represent random variables and shaded nodes are observed during learning.

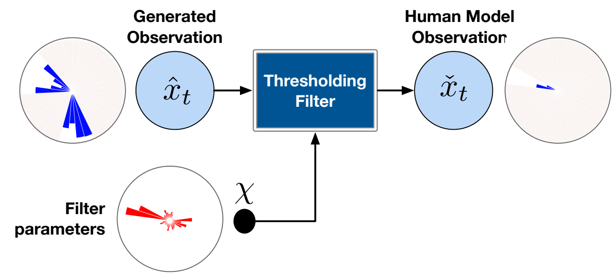


Fig. 3: An Example Thresholding Filter as a Perceptual Implant. The (generated) range observations (blue bars) are passed through a thresholding filter to eliminate observations that are beyond a parameterized distance (red bars) in each segment from the human.

simulates possible futures using its internal models, and plans communication to maximize task rewards while minimizing communication costs.

## III. MODEL IMPLANTS FOR RAPID REFLECTIVE OTHER-AGENT REASONING/LEARNING (MIRROR)

Our problem setting is one of assistance: a (human) user is acting in a partially-observable environment to maximize rewards. The assistive robot's goal is to help the user achieve their objective. The robot may receive different observations from the environment and can modify the user's observations to provide additional information. We seek to derive an effective assistant. At a high level, we will imbue the robot with a structured model of the human that can be adapted with data. After learning, the robot plans using its own internal (self) model and the human model to communicate valuable information. We first detail the robot's underlying self-model, then describe the human model (specifically the implants), and finally, how communication can be achieved using both the self and human models.

### A. Self Model: Multi-Modal Latent State-Space Model

In MIRROR, the robot's self-model is a multi-modal state-space model (MSSM) [31] (Fig. 2). Intuitively, the MSSM models an agent that is sequentially taking actions in a world and receiving rewards and multi-modal observations. The observation $x_t^m$ at time $t$ for $m = 1, \ldots, M$ sensory modalities are generated from the latent state $z_t$. The model assumes Markovian transitions where the next state is conditioned upon the current state and the action $a_t$ taken by the agent. Upon taking an action, the agent receives reward $r_t$. In general, the reward can also be conditioned upon the action and next state. The MSSM is trained via amortized variational inference and we used Soft-Actor Critic (SAC) [32] to train the robot policy.

### B. Learning Human Models via MIRROR

In MIRROR, the human model is identical to the robot's self-model *except* for implants that are injected to change the model's behavior. We use the superscript H to refer to the human model.

**Model Implants.** Once the self-model is trained, we augment it with implanted functions $h(\cdot)$. In this work, we examine two implant classes:

- **Perceptual implants** model how humans perceive the world by changing the observation $x_t$. In our experiments, we use a threshold mask/filter implant that models that the human can only see objects within a certain range (Fig. 3)
- **Policy implants** model how the human acts differently from the robot by changing the policy. Simple implants can simply add noise to increase policy entropy. In this work, we add a state-dependent residual $\delta(z_t^{\mathrm{H}})$ term to the parameters of the policy distribution $\pi$. As an example, for the Gaussian policies $\mathcal{N}\left(\mu(z_t^{\mathrm{H}}), \sigma^2(z_t^{\mathrm{H}})\right)$ used in our CARLA experiments, we add $\delta(z_t^{\mathrm{H}})$ to the mean of the Gaussian such that the augmented policy distribution follows $\mathcal{N}\left(\mu(z_t^{\mathrm{H}}) + \delta(z_t^{\mathrm{H}}), \sigma^2(z_t^{\mathrm{H}})\right)$.

The above provides a flavor of what is possible and is not exhaustive. For example, to model humans who are slow to change their beliefs, we could implant a low-pass filter $z_{t+1}^{\mathrm{H}} = \alpha z_t + (1-\alpha) f(z_t, x_t, a_t)$ where $\alpha$ is a learnable parameter. We leave exploration of other implants to future work.

**Learning Implant Parameters.** Given an implant $h_\chi$ parameterized by $\chi$, we can learn $\chi$ by minimizing the following loss given data:

$$\arg\min_{\chi} \mathcal{L}(\chi) = -\sum_{t=1}^{T} \mathbb{E}_{z_t \sim p(z_t | x_{1:t}^{1:M}, a_{1:t-1}, \chi)} \left[\log \pi(a_t | z_t)\right] - \lambda \log p(\chi) \quad (1)$$

where $\lambda$ is a regularization hyperparameter that controls the strength of an optional prior. Intuitively, this loss optimizes the likelihood of observing a human's actions given the self-model and the implant parameters. In our work, we approximate $p(z_t | x_{1:t}^{1:M}, a_{1:t-1})$ using the learnt inference network and perform stochastic gradient descent by sampling $z_t$. Note that *only* the implant parameters are modified when learning from human data.

### C. *Human-Robot Communication with* MIRROR

Next, we turn our attention to how the implanted self-model can be used for human-robot communication. The key idea is to plan using both learned models; we couple the robot's self-model and the human model together via generated robot observations and communication actions, and predicted human actions (Fig. 1.B.). We use a model-based approach, the robot forward simulate trajectories by using its self-model and human model and optimize communication to maximize task rewards while minimizing costs:

$$\arg\max_{\omega_{0:T}} J = \mathrm{E}_{z_{0:T}|\omega_{0:T}}\left[\sum_{t=0}^{T}\gamma^t\left(r(z_t) - C(\omega_t)\right)\right] \quad (2)$$

where $\omega_{0:T}$ is the parameters of our communication filter for time-steps 0 to $T$, $\gamma$ is the discount factor, $r$ is the task reward function[1], and $C$ is the cost function. The expectation is taken with respect to the trajectories $\tau$ under the models and filter parameters $\omega_{0:T}$. In our work, we use the cross-entropy (CE) method [33], and re-plan at each time-step. Note that real observations are obtained after each step, which are used to update the beliefs of the robot and human models.

## IV. HUMAN-SUBJECT EXPERIMENTS

In this section, we report on experiments designed to test whether MIRROR is able to provide useful information to human users in a more realistic setting. We use CARLA [10], a modern driving simulator (Fig 4). The CARLA environment has continuous state and action spaces, with realistic dynamics and visuals. Our main hypotheses were that (**H1**) planning with implanted self-models would yield more helpful communication than planning with behavioral cloning models and (**H2**) planning to optimize task rewards and communication costs would lead to less redundant communication compared to belief matching. Our study was approved by our institution's ethical review board.

### A. *Experimental Setup*

**Task Description.** Participants interacted with our assistive driving agents in a highway driving task under dense fog (see Fig. 4.C and 4.D for a comparison between clear and adverse weather conditions). The goal was to drive along a stretch of CARLA's Town04's two-lane carriageway from the starting position to the destination, while navigating through a normal highway traffic. The other vehicles may slow down or speed up, and due to poor visibility, would not be able to avoid the participant's vehicle. The participant had to actively avoid other vehicles along the way.

**Assistive Communication.** The car is equipped with a semantic LIDAR and a driving assistant that can provide both visual and verbal cues. Specifically, the agent could highlight selected vehicles through visual bounding boxes and/or provide informative speech (as previously shown in Fig.

[1]In principle, this task reward function may differ from the reward function that the model was trained with, but we leave such experiments to future work.

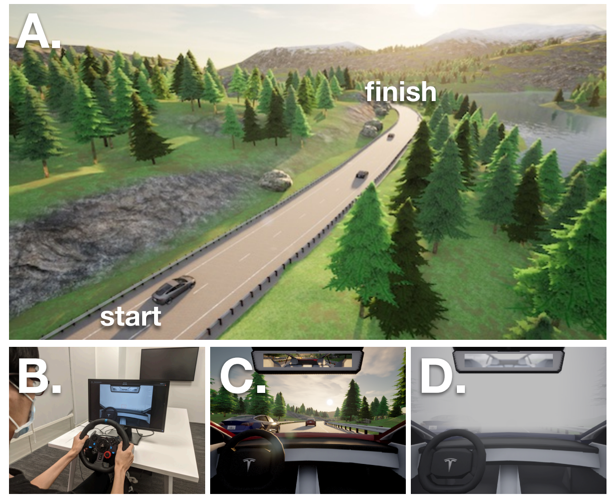


Fig. 4: CARLA Experiment Setup. (**A**) The stretch of highway that participants drove along. (**B**) Participants drove the simulated car using a steering wheel with accelerator and brake pedals. (**C**) and (**D**) show the difference in visibility in clear and foggy weather. Both cars are visible in the clear setting. In the fog setting, the car on the left is visible, but the car in the front can barely be seen.

1.A). The system is capable of informing participants when a car approaches or slows down, as well as when no cars are detected in a specific direction.

**Compared Methods.** We trained MIRROR with both perceptual and policy implants. We compared four conditions:

- No Communication (NC): the human receives no assistance.
- Behavioral Cloning (BC): the human model is trained using behavioral cloning.
- MIRROR: our MIRROR method that first trains a self-model via RL, then learns the implants from demonstrations, and plans communication.
- MIRROR-KL: a MIRROR variant that does not plan but minimizes the KL-Divergence between the human's mental state and the robot's belief, similar to [3].

**Participants.** A total of 21 participants (mean age = 23.2, 10 females) were recruited from the university community. The experiment was designed to be within-subjects with all 21 participants in each condition.

**Procedure.** Participants entered the lab and were briefed about the task. They then engaged in two practice trials; the first trial involved driving freely along the highway in clear weather conditions, and the second trial involved three rounds of the driving task under dense fog conditions without any assistive communication. Thereafter, they performed a total of 24 rounds, with the first 6 rounds without any assistive communication, followed by 18 rounds with three different agents (six rounds per agent). The data from the first 6 rounds were used to train the models and the order of the three agents was counterbalanced. Participants could choose to take a one minute rest after every 6-th round to reduce fatigue.

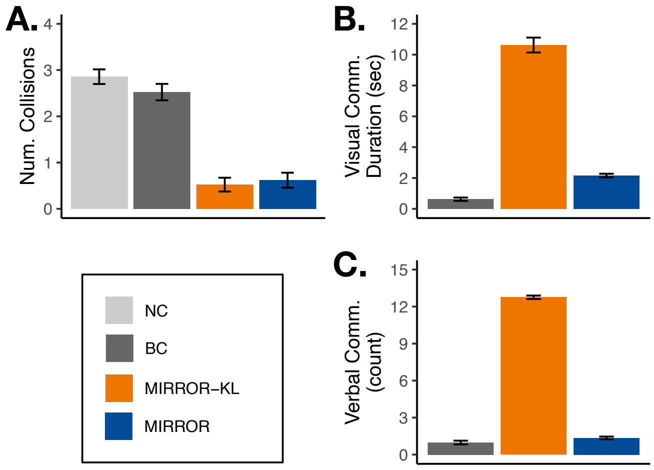


Fig. 5: Objective Measures. Error bars indicate one standard error. (**A**) Number of Collisions. (**B**) Amount of Visual Communication. (**C**) Amount of Verbal Communication (**D**). MIRROR results in fewer collisions compared to behavioral cloning (BC) and no communication (NC) conditions, but with far less communication required compared to belief matching (MIRROR-KL).

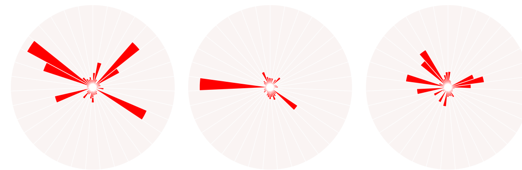

Fig. 6: Samples of Learned Perceptual Implants (threshold filters) for the CARLA experiment. Top is the front of the vehicle. Length of red bars indicate visibility distance. The implants indicate the human was not able to see far ahead or in the rear but could see cars at the side. Compare against Fig. 4.C and 4.D.

**Dependent Measures.** We use both objective and subjective measures to evaluate agent performance. Objective measures comprised the number of collisions with other vehicles/environment, and the amount of communication (e.g., the number of speech utterances). We also collected a range of subjective measures to ascertain cognitive load, communication properties (helpfulness, redundancy, timeliness, modality selection), and trust after interaction with each agent.

### *B. Results and Analysis*

In brief, the results support both hypotheses. We compared the methods along both objective and subjective dimensions using a repeated measures one-way ANOVA, followed by selected pairwise $t$-tests with adjusted-$\alpha = 0.0167$ using Bonferroni correction.

Fig. 5.A shows that participants experienced significantly fewer collisions when interacting with MIRROR ($F_{3,60} = 47.977$, $p < 0.001$; MIRROR vs BC: $t(9.5) = 7.683$, $p < 0.001$). Subjectively, participants found MIRROR provided information that was more helpful and timely compared to BC, and were also more comfortable with the communication modality chosen ($p < 0.005$ across the measures and pairwise tests). Participants also trusted the MIRROR agent more than BC ($F_{3,60} = 47.730$, $p < 0.001$; MIRROR vs BC: $t(9.5) = -12.526$, $p < 0.001$). The overall Raw-TLX scores indicated that the participants felt less mentally burdened when they interacted with MIRROR ($F_{3,60} = 10.071$, $p < 0.001$; MIRROR vs BC: $t(9.5) = 6.658$, $p < 0.001$).

Taken together, both objective and subjective evidence strongly support hypothesis **H1**. This finding is corroborated by participant survey responses; they shared that MIRROR "*conveys critical information at a good timing*" in contrast to BC, which they felt "*is not helping me at all*" and "*doesn't inform me about the cars that are approaching from the rear*". Qualitatively, we found planning with the BC human model to be inaccurate; the BC model would quickly overfit, which led to poor communication. In contrast, the MIRROR implant models resulted in better communication; Fig. 6 shows the perceptual implants learned by MIRROR to well approximate what the human could see, even with a small amount of training data (six demonstrations).

Next, we turn our attention to **H2**, i.e., whether planning with task rewards and communication costs reduced redundant communication relative to belief matching. Figs. 5.B and 5.C show how long each agent highlighted cars and the number of times they verbally alerted the driver, respectively. MIRROR-KL tended to be overly communicative; it provided more than 5 times more visual and verbal communication compared to MIRROR without significant benefits in terms of task performance. Subjectively, participants rated MIRROR-KL to provide more redundant information compared to MIRROR ($t(9.5) = -11.411$, $p < 0.001$). In their survey responses, participants wrote that the MIRROR-KL agent "*is too talkative*", "*told me a lot of useless information*" and "*is distracting and annoying*". In comparison, they found MIRROR to be "*straight to the point.*" Recall that the only difference between the two is the objective function; MIRROR-KL tries to align belief distributions, regardless of whether the alignment leads to better task accomplishment. We observed that MIRROR-KL would communicate verbally even when there was no need to, e.g., it would repeatedly tell participants that "*there is no car in the rear*".

When asked which agent they were most comfortable with, a majority of the participants (14 out of 21) selected MIRROR. The remaining participants picked MIRROR-KL. When asked about which agent they would prefer for long-distance driving (above 1 hour), the number of participants selecting MIRROR increased to 18. No participant selected BC. Interestingly, some participants preferred MIRROR-KL's talkative nature, with one participant stating that they "*felt annoyed and safe at the same time*". A few participants found MIRROR to be too quiet: "*I have some confidence that it works but I'm not entirely sure because it is quieter*". These responses suggest individual preferences for information/reassurance and differing trust in the system. How we can incorporate these aspects within MIRROR would make for interesting future work.

## V. Conclusion

In summary, we present MIRROR, a framework for learning human models using deep self-models for initial structure, along with a planning-based communication approach that couples the human model with learned world model. Experiments show MIRROR to be effective, outperforming behavioral cloning and belief-matching. The results show that bootstrapping human model learning with latent-variable models learnt during reinforcement learning leads to generalizable models that are more useful for interaction planning. More broadly, we consider MIRROR to be a step towards more data-efficient human models for human-robot interaction. The key idea examined in this work—learning differences from the robot self-model—can be potentially be applied to general human-robot collaboration. Compared to existing work on human models, MIRROR embodies an alternative paradigm that "front-loads" the learning the environmental dynamics and task structure, and thus, offers savings both in terms of sample complexity and computation when learning from human demonstrations. We believe that building upon MIRROR forms a compelling pathway towards robots that can interact fluently with humans.

## References


[1] A. Tabrez, S. Agrawal, and B. Hayes, "Explanation-based reward coaching to improve human performance via reinforcement learning," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2019, pp. 249–257.

[2] M. C. Buehler and T. H. Weisswange, "Theory of mind based communication for human agent cooperation," in *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*, 2020, pp. 1–6.

[3] S. Reddy, S. Levine, and A. D. Dragan, "Assisted perception: optimizing observations to communicate state," in *Conference on Robot Learning (CoRL 2020)*, 2020.

[4] F. H. Allport, *Chapter 13: Social Attitudes and Social Consciousness*, ser. Social Psychology. Houghton Mifflin Company, 1924.

[5] J. I. Krueger, T. E. DiDonato, and D. Freestone, "Social projection can solve social dilemmas," *Psychological Inquiry*, vol. 23, no. 1, pp. 1–27, 2012.

[6] J. I. Krueger, "Social projection as a source of cooperation," *Current Directions in Psychological Science*, vol. 22, no. 4, pp. 289–294, 2013.

[7] M. Carroll, R. Shah, M. K. Ho, T. Griffiths, S. Seshia, P. Abbeel, and A. Dragan, "On the utility of learning about humans for human-ai coordination," *Advances in Neural Information Processing Systems*, vol. 32, pp. 5174–5185, 2019.

[8] S. Bekkali, G. J. Youssef, P. H. Donaldson, N. Albein-Urios, C. Hyde, and P. G. Enticott, "Is the putative mirror neuron system associated with empathy? a systematic review and meta-analysis," *Neuropsychology review*, vol. 31, no. 1, pp. 14–57, 2021.

[9] G. Rizzolatti and L. Craighero, "The mirror-neuron system," *Annu. Rev. Neurosci.*, vol. 27, pp. 169–192, 2004.

[10] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proceedings of the 1st Annual Conference on Robot Learning*, 2017, pp. 1–16.

[11] A. Tabrez, M. B. Luebbers, and B. Hayes, "A Survey of Mental Modeling Techniques in Human–Robot Teaming," *Current Robotics Reports*, vol. 1, no. 4, pp. 259–267, 2020.

[12] R. Choudhury, G. Swamy, D. Hadfield-Menell, and A. D. Dragan, "On the utility of model learning in hri," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2019, pp. 317–325.

[13] N. Mavridis, "A review of verbal and non-verbal human-robot interactive communication," *Robotics and Autonomous Systems*, vol. 63, no. P1, pp. 22–35, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.robot.2014.09.031

[14] C. Schaff and M. R. Walter, "Residual policy learning for shared autonomy," in *Proceedings of Robotics: Science and Systems (RSS)*, 2020.

[15] C. L. Baker, J. Jara-Ettinger, R. Saxe, and J. B. Tenenbaum, "Rational quantitative attribution of beliefs, desires and percepts in human mentalizing," *Nature Human Behaviour*, vol. 1, no. 4, pp. 1–10, 2017.

[16] M. K. Ho and T. L. Griffiths, "Cognitive science as a source of forward and inverse models of human decisions for robotics and control," *arXiv preprint arXiv:2109.00127*, 2021.

[17] H. Soh, Y. Xie, M. Chen, and D. Hsu, "Multi-task trust transfer for human–robot interaction," *The International Journal of Robotics Research*, vol. 39, no. 2-3, pp. 233–249, 2020.

[18] M. Kwon, E. Biyik, A. Talati, K. Bhasin, D. P. Losey, and D. Sadigh, "When humans aren't optimal: Robots that collaborate with risk-aware humans," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 43–52.

[19] S. Reddy, A. D. Dragan, and S. Levine, "Where do you think you're going?: Inferring beliefs about dynamics from behavior," in *NeurIPS*, 2018.

[20] I. Horswill, "Polly: A vision-based artificial agent," in *AAAI*, 1993, pp. 824–829.

[21] J. Buhmann, W. Burgard, A. B. Cremers, D. Fox, T. Hofmann, F. E. Schneider, J. Strikos, and S. Thrun, "The mobile robot rhino," *Ai Magazine*, vol. 16, no. 2, pp. 31–31, 1995.

[22] A. D. Dragan, K. C. Lee, and S. S. Srinivasa, "Legibility and predictability of robot motion," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2013, pp. 301–308.

[23] A. D. Dragan, S. Bauman, J. Forlizzi, and S. S. Srinivasa, "Effects of robot motion on human-robot collaboration," in *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2015, pp. 51–58.

[24] B. Busch, J. Grizou, M. Lopes, and F. Stulp, "Learning legible motion from human–robot interactions," *International Journal of Social Robotics*, vol. 9, no. 5, pp. 765–779, 2017.

[25] T. Kollar, S. Tellex, D. Roy, and N. Roy, "Toward understanding natural language directions," in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2010, pp. 259–266.

[26] J. Thomason, M. Murray, M. Cakmak, and L. Zettlemoyer, "Vision-and-dialog navigation," in *Conference on Robot Learning*. PMLR, 2020, pp. 394–406.

[27] V. V. Unhelkar, S. Li, and J. A. Shah, "Decision-making for bidirectional communication in sequential human-robot collaborative tasks," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 329–341.

[28] D. Das, S. Banerjee, and S. Chernova, "Explainable ai for robot failures: Generating explanations that improve user assistance in fault recovery," in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 2021, pp. 351–360.

[29] J. Lee, J. Fong, B. C. Kok, and H. Soh, "Getting to know one another: Calibrating intent, capabilities and trust for human-robot collaboration," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 6296–6303.

[30] E. OhnBar, K. Kitani, and C. Asakawa, "Personalized dynamics models for adaptive assistive navigation systems," in *Conference on Robot Learning*. PMLR, 2018, pp. 16–39.

[31] K. Chen, Y. Lee, and H. Soh, "Multi-modal mutual information (mummi) training for robust self-supervised deep reinforcement learning," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2021.

[32] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *International conference on machine learning*. PMLR, 2018, pp. 1861–1870.

[33] R. Rubinstein, "The cross-entropy method for combinatorial and continuous optimization," *Methodology and computing in applied probability*, vol. 1, no. 2, pp. 127–190, 1999.