Predictability Awareness For Efficient and Robust Multi-Agent Coordination

Roman Chiva Gil
Delft University Of Technology
Delft, Netherlands
R.ChivaGil@student.tudelft.nl

Daniel Jarne Ornia*†
University of Oxford
Oxford, United Kingdom
daniel.jarneornia@cs.ox.ac.uk

Khaled A. Mustafa*
Delft University Of Technology
Delft, Netherlands
k.a.mustafa@tudelft.nl

Javier Alonso Mora
Delft University Of Technology
Delft, Netherlands
j.alonsomora@tudelft.nl

ABSTRACT

To safely and efficiently solve motion planning problems in multi-agent settings, most approaches attempt to solve a joint optimization that explicitly accounts for the responses triggered in other agents. This often results in solutions with an exponential computational complexity, making these methods intractable for complex scenarios with many agents. While sequential predict-and-plan approaches are more scalable, they tend to perform poorly in highly interactive environments. This paper proposes a method to improve the interactive capabilities of sequential predict-and-plan methods in multi-agent navigation problems by introducing predictability as an optimization objective. We interpret predictability through the use of general prediction models, by allowing agents to predict themselves and estimate how they align with these external predictions. We formally introduce this behavior through the free-energy of the system, which reduces (under appropriate bounds) to the Kullback-Leibler divergence between plan and prediction, and use this as a penalty for unpredictable trajectories. The proposed interpretation of predictability allows agents to more robustly leverage prediction models, and fosters a 'soft social convention' that accelerates agreement on coordination strategies without the need of explicit high level control or communication. We show how this predictability-aware planning leads to lower-cost trajectories and reduces planning effort in a set of multi-robot problems, including autonomous driving experiments with human driver data, where we show that the benefits of considering predictability apply even when only the ego-agent uses this strategy. The code and experiment videos can be found in the following page: https://romanchiva.github.io/PAProjectPage/

KEYWORDS

Multi-Agent Systems, Motion Planning, Autonomous Navigation, Coordination, Prediction Models

1 INTRODUCTION

Many modern robotics applications involve autonomous agents navigating multi-agent environments where they will be required to interact with humans and other robots without full knowledge or extensive communication capabilities [34]. This involves planning trajectories in a complex system governed by a mix of rational and non-rational, stochastic and possibly game theoretic behaviors. To achieve safe and efficient interactions, agents need to reason about each other and coordinate. However, this poses critical challenges due to the high uncertainty associated with estimating other agents's objectives [16] and a computational complexity that renders problems intractable for more than a handful of agents.

Receding Horizon Trajectory Optimization allows for flexible and anticipative planning while ensuring compliance with *e.g.* safety constraints in multi-agent navigation problems. However, planning a trajectory that explicitly accounts for interactions among agents generally requires solving a joint optimization problem. A variety of joint planning methods can be found in literature, *e.g.* [10, 22], of which game theoretic approaches best capturing agent interaction complexities [34]. By modeling other agents as rational actors, game theoretic approaches cast the joint optimization as a constrained dynamic game and seek to find equilibrium solutions. Although this often results in stable and coordinated interactions [11, 16], game theoretic approaches suffer from the curse of dimensionality, as the planning complexity grows exponentially with the number of agents [31]. Additionally, modeling other agents as rational is a strong assumption which will not hold in practice, especially when interacting with human agents [4, 12].

Alternatively, predict-and-plan approaches scale well with number of agents, however they tend to perform poorly in interactive environments. By separating prediction and planning, the problem simplifies to a single-agent collision avoidance problem with dynamic obstacles [5, 13]. The accuracy of the prediction model limits how well agents can coordinate. A system of interacting agents is highly complex, making it difficult to predict the diversity of possible futures, especially when considering interactions. This can lead to ambiguous predictions, making agents unable to anticipate their environment, and thus have to re-plan more often or engage in riskier behaviors [34].

Ideally, every agent in the environment would be able to accurately anticipate surrounding agents' future trajectories allowing

*Indicates equal contribution, This research is supported by funding from the Dutch Research Council NWO-NWA, within the "Acting under Uncertainty" (ACT) project (Grant No.NWA.1292.19.298), † Work done partially while at Delft University of Technology. Author acknowledges partial support from UKRI grant EP/W002949/1. .

*Equal contribution.
This research is supported by funding from the Dutch Research Council NWO-NWA, within the "Acting under Uncertainty" (ACT) project (Grant No.NWA.1292.19.298).
†Work done partially while at Delft University of Technology. Author acknowledges partial support from UKRI grant EP/W002949/1

for efficient and safe interaction. Sequential planning agents use prediction models to avoid collisions with others, however, this fails to acknowledge that surrounding agents also hold predictions about the ego-agent, and plan their trajectory based on these predictions. Unless the optimal avoidance strategy falls within the range of predicted behaviors, other agents will react to the unexpected avoidance strategy by modifying their own trajectory. To mitigate this issue, we propose the following: in the same way a prediction model is used to predict other agents, the ego-agent can use it to approximate how other agents expect it to behave. This information can be used in planning to introduce a penalty for trajectories other agents will find surprising, bringing the optimal trajectory closer to the expectation surrounding agents hold. Accounting for predictability in this way mirrors the principle of free-energy minimization in active inference [30] (and control systems [32]), where an agent not only seeks to maximize reward but also aims to minimize the discrepancy between some prediction model and observations. In multi-agent interactions [24], agents hold probabilistic beliefs about the behavior of others, and the accuracy of these beliefs is directly influenced by the agent's own actions. By minimizing free energy, the agent balances actions that reduce uncertainty and confirm its internal model of the world with those that maximize reward. This approach ensures that the agent's behavior is not only goal-directed but also aligned with maintaining coherent and accurate beliefs about the surrounding agents.

1.1 Contribution

We explore how sequential planning agents can improve their coordination capabilities by accounting for the predictability of their planned trajectories. When a group agents accounts for predictability, they are able to foster a 'soft social convention' dictated by the prediction model which results in a decrease of uncertainty about the environment for all agents in the group. This helps agents resolve coordination problems without having to explicitly model interactions. Formally, the contribution of this paper is threefold:

(1) We exploit ideas on free-energy to formulate a cost function that uses feedback from a prediction model to include predictability as an objective and analyze how this cost function can be integrated with a planner.
(2) We provide results showing how our predictability awareness mechanism leads to 'soft social conventions' forming-based interaction strategies encoded in prediction models for multi-robot navigation problems. This allows agents to achieve smoother coordination by improving the effectiveness of prediction models in interactive environments.
(3) Accounting for predictability causes agents to adopt *social norms* and pro-social behaviors encoded in learned prediction models, allowing to more closely mimic experts' behaviors without needing cost function learning. We provide evidence for these behaviors in an experiment where an agent interacts with human drivers in scenarios from the Waymo Open Motion Dataset.

2 RELATED WORK

Integration of Prediction Models and Planners. Trajectory prediction has significantly advanced in recent years, particularly with the development of transformer-based generative models capable of producing interaction-aware joint trajectory predictions, *e.g.* [8, 9, 21]. While these models show impressive performance in open-loop evaluations, integrating them with planners in highly interactive settings remains challenging [20]. Effective interactive planning often necessitates joint prediction and planning. Additionally, the planner often requires some form of learned cost function [23]. Otherwise, if the behavior of the expert significantly differs from the expert in the training data, this will throw the model out of distribution yielding low quality predictions.

Many studies have focused on developing ego-conditioned prediction models [27]; however, their integration with planners faces obstacles primarily due to computational complexity. For instance, in [22] Tree Policy Planning (TPP) has been employed to generate an initial set of partial trajectories, which condition the prediction model and create a scenario tree. This tree is evaluated using a cost function combining designed and learned features to identify and expand promising scenarios, efficiently allocating computational resources. A novel approach by [10] leverages unconditioned prediction models to provide initial estimates of other agents' trajectories, capitalizing on the models' ability to predict general intentions accurately while acknowledging their limitations in capturing short-term interaction details. This approach optimizes the ego and agent trajectories together, minimizing disturbances from the initial agent paths and utilizing homotopy classes to ensure diversity and avoid local minima. Instead of conditional prediction models, some methods develop fully differentiable stacks [23, 25] enabling gradient backpropagation through the planner, which allows for combined prediction model fine-tuning and cost function learning aligned with expert behavior in the training data. While avoiding the joint optimization, our approach links prediction and planning without the need for retraining or fine-tuning by including a term in the cost function that helps guide the agent's behavior to not compromise its predictions. This allows for maintaining flexibility in selecting prediction models and planner combinations while being compute-efficient.

Predictability and Legibility of Motion. In the field of Human-Robot Interaction, legibility and predictability of motion have been studied to improve coordination by designing agent behaviors that clearly communicate intention and avoid surprising observers [15]. Often both objectives overlap [3]. Traditional formulations of this problem are not well suited for receding horizon applications as they optimize over complete trajectories and rely on utility-based analytical models of observer expectations [14]. Additionally, the observer is modeled as inactive, thus having no influence on the planning agent. This assumption breaks down in multi-agent navigation where the observer and the agent share the workspace and influence each other. Several works have explored the adaptation of these concepts to an interactive multi-agent context. In single agent RL settings, the impact of predictability objectives has been recently studied in [28]. In multi-agent scenarios, [2] define dynamic goal regions around neighboring agents and optimize for reduced uncertainty about the collision avoidance strategy. [19]

show how increasing action penalties at later horizon steps causes agents to more rapidly demonstrate their avoidance strategy. This accelerates intent inference giving agents better anticipation. [6] defines hand-crafted legibility costs for planning in highway driving. These methods are often designed to target a specific type of observer model. In contrast, our approach minimizes a predictability surrogate that allows modeling the observers with an arbitrary prediction model choice.

3 TRAJECTORY PLANNING

The general optimization problem for a single-agent in stochastic motion planning can be formulated as follows:

$$\min_{\boldsymbol{u}\in\mathbb{U},\boldsymbol{x}\in\mathbb{X}} \quad \sum_{k=0}^{K-1} J_k(\boldsymbol{x}_k,\boldsymbol{u}_k) + J_K(\boldsymbol{x}_K) \tag{1a}$$

$$\text{s.t.} \quad \boldsymbol{x}_0 = \boldsymbol{x}_{\text{init}}, \tag{1b}$$

$$\boldsymbol{x}_{k+1} = f(\boldsymbol{x}_k,\boldsymbol{u}_k), \quad k = 0,...,K-1 \tag{1c}$$

$$\mathbb{P}\left[C(\boldsymbol{x}_k,\boldsymbol{\delta}_k^o),\forall o\right] \geq 1-\epsilon_k,\forall k, \tag{1d}$$

where $\boldsymbol{u} = \{\boldsymbol{u}_0,...,\boldsymbol{u}_K\} \in \mathbb{U}$ are the system inputs subject to input constraints, $\boldsymbol{x}_k \in \mathbb{X}$ denotes the states of the robot, $f(\cdot)$ corresponds to the nonlinear system's dynamics, $J_k(\boldsymbol{x}_k,\boldsymbol{u}_k) \geq 0$ is the cost function specifying performance metrics, and K is the length of the planning horizon. In this formulation, $C(\cdot)$ is the collision avoidance constraint, and $\boldsymbol{\delta}_k^o$ is the uncertain position of obstacle o at stage k obtained through a prediction model $\mathcal{P}(\mathbf{X})$ that takes into account the concatenated states of all agents in the scene. The chance constraint in Eq. (1d), guarantees that the probability that the robot collides with the dynamic obstacle is below a specified threshold ϵ_k.

In a game theoretic setting where all agents are controlled by a centralized planner, the problem reduces to solving a joint optimization program over all agents and all possible trajectories, such that from the set of joint trajectories that satisfy the constraints, the agents execute the optimal ones. This naturally carries high computational complexity, access to some centralized controller, and full information assumptions. Consider instead the case where N (interactive) agents solve the optimization problem (1) independently and use model $\mathcal{P}(\mathbf{X})$ to predict each other (and thus estimate the probabilities of constraint satisfaction). Agents can then query the prediction model to observe the predictions others have about them. Our method reduces to the following intuition: Agents can use this information to shift their behaviors towards the distribution coming out of the prediction model. This 'closes the loop' on prediction errors, intuitively improving the planning problem in two ways. First, inducing implicit decentralized coordination: an ideal situation is one where all agents act following the model-predicted distribution, and this distribution perfectly optimizes the cost of each agent. Second, it 'robustifies' the prediction model *a posteriori*: once the model has been trained on offline data, agents actively shift their plans towards the predicted distributions, collectively reducing prediction errors and widening the space of suitable prediction models for a given problem.

4 PROPOSED METHOD: FREE ENERGY AS A PREDICTABILITY SURROGATE

4.1 Derivation of a predictability aware cost function

Our objective is to design a framework that allows agents to trade off predictability with progress toward the goal. If we define an agent's optimal trajectory distribution as Q^*, in the best-case scenario, an agent's optimal trajectory distribution aligns with the predictions held by other agents. This alignment allows the agent to minimize its own cost while avoiding any disruption or interference with the trajectories of surrounding agents. In this case, no trade-off needs to be performed, however, deviations from this ideal scenario are to be expected. To formalize this as a planning objective, agents should seek to minimize the cost of trajectories sampled from their corresponding prediction in $\mathcal{P}(\mathbf{X})$. Drawing inspiration from the path integral control derivation in [36], we begin by defining the free energy of a trajectory distribution:

$$\mathcal{F}(S,\mathcal{P},\lambda) = -\lambda\log(\mathbb{E}_{\boldsymbol{x}\sim\mathcal{P}}[\exp(-\frac{1}{\lambda}S(\boldsymbol{x}))]),$$

where S is a state cost function that *represents some (trajectory planning) objective*, $\mathcal{P}$ denotes a prediction distribution, x is a trajectory sampled from $\mathcal{P}$, and λ represents the inverse temperature controlling the strictness of the efficiency criterion. This control theoretic free energy can be interpreted as a measure of how efficient a prediction distribution is at minimizing cost S. The free energy is minimized by pushing $\mathcal{P}$ as close as possible to Q^*.

The free energy as defined so far is a function of prediction distribution $\mathcal{P}$, however, agents won't plan trajectories by sampling from $\mathcal{P}$. Instead, we define Q as a trajectory distribution an agent has control over. Let the states $\boldsymbol{x} = \{\boldsymbol{x}_0,...,\boldsymbol{x}_K\}$, which the ego-agent occupies along its planned trajectory $\tau_{0,K}$, be represented as narrow Gaussians $q(\boldsymbol{x}_k)$ with mean $\boldsymbol{x}_k$ covariance Σ_k:

$$\begin{aligned}\tau_{0:K} &= \{q(\boldsymbol{x}_k)\}_{k=0}^K,\\ q(\boldsymbol{x}_k) &= \mathcal{N}(\boldsymbol{x}_k,\Sigma_k).\end{aligned} \tag{2}$$

By applying an expectation switch, these distributions can be incorporated into the free energy definition, making it a function of the agent's plan,

$$\mathcal{F}(S,\mathcal{P},\lambda) = -\lambda\log(\mathbb{E}_{\boldsymbol{x}\sim Q}[\exp(-\frac{1}{\lambda}S(\boldsymbol{x}))\frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}]), \tag{3}$$

where p is the density function of the prediction. By concavity of the logarithm and Jensen's inequality,

$$\mathcal{F}(S,\mathcal{P},\lambda) \leq -\lambda\mathbb{E}_{\boldsymbol{x}\sim Q}[\log(\exp(-\frac{1}{\lambda}S(\boldsymbol{x}))) + \log(\frac{p(\boldsymbol{x})}{q(\boldsymbol{x})})].$$

Finally, using the definition of Kullback-Leibler Divergence and simplifying,

$$\mathcal{F}(S,\mathcal{P},\lambda) \leq \mathbb{E}_{\boldsymbol{x}\sim Q}[S(\boldsymbol{x})] + \lambda\mathbb{KL}(q(\boldsymbol{x})||p(\boldsymbol{x})), \tag{4}$$

where $\mathbb{KL}$ denotes the KL-Divergence. The right-hand side provides an upper bound on the free energy, and one can minimize this instead of the free energy. It resembles a standard control objective, and the terms allow for good conceptual understanding of the effect they have: A **Performance Cost** and **Predictability**

Cost respectively, which penalizes agents for acting unpredictably. Using this newly found expression as a stage cost, we can craft the following cost function as a stage cost for a planning problem:

$$J(\tau_{0:K}) = \sum_{k=0}^{K} J_k(\boldsymbol{x}_k, \boldsymbol{u}_k) + \lambda \mathbb{KL}(q(\boldsymbol{x}_k)||p(\boldsymbol{x}_k)),$$

where we implicitly assume J_k to be composed by some state cost S and some control action cost. Minimizing this cost function allows agents to trade off predictability and progress toward the goal by means of the free energy, and λ can be selected to control how much weight is assigned to predictability during planning.

Remark 1. *We can emphasize now the intuition behind using the free energy as a way of incorporating predictability into optimal control. Eq.* (4) *is minimized precisely when $Q = Q^* = \mathcal{P}$. That is, the trajectory distribution executed is exactly the optimal cost trajectory distribution, and this matches the predicted distribution. Under this condition, the agent is behaving without surprising external observers and simultaneously obtaining optimal cost in its objective.*

4.2 Integration with a Planner and Practicalities

The KL-Divergence expression only has closed form solutions for a restricted set of distributions, thus to accommodate arbitrary distributions, the KL divergence term will often need to be evaluated through sampling with Q the candidate trajectory distribution and P the prediction distribution from $\mathbb{KL}(\mathcal{P}\|Q) = \mathbb{E}_{\boldsymbol{x}\sim P}\left[\log \frac{\mathcal{P}(\boldsymbol{x})}{Q(\boldsymbol{x})}\right]$. Since sampling is required to evaluate the cost function, this could render the use of gradient based MPC unfeasible for real time planning, additionally prediction distributions $Q(\boldsymbol{x})$ may not always be differentiable. We find it is more practical to rely on sampling based MPC approaches, as they don't require a differentiable cost function and computations can be easily parallelized to handle large numbers of samples even when it is computationally expensive to evaluate the cost function. In our experiments, Section 5, we rely on an Model Predictive Path Integral (MPPI) control method [36].

Another consideration is that predictions about an agent's future are updated as new observations are received. For this reason, it is most effective to focus on early horizon time-steps when evaluating a plan's predictability. Thus we propose to discount the predictability cost along the horizon with factor γ to account for uncertainty about future predictions:

$$J(\tau_{0:K}) = \sum_{k=0}^{K} J_k(\boldsymbol{x}_k, \boldsymbol{u}_k) + \gamma^k \lambda \mathbb{KL}(q(\boldsymbol{x}_k)||p(\boldsymbol{x}_k)). \tag{5}$$

Remark 2. *It should be noted that our method is agnostic to the choice of the planner. However, in case MPPI is used as the planner, similar to [33], our approach can be interpreted as leveraging the distribution of the prediction model as an ancillary controller to influence the MPPI sampling process.*

5 EXPERIMENTS

We present here the experiments carried out to validate our method. The first experiment investigates how accounting for predictability affects an individual agent's behavior, comparing the results with other observer-aware planning approaches. The second experiment examines the impact of predictability within a group of agents, focusing on swapping tasks in an open environment to give insight without external environmental influences. In the third experiment, we explore a practical driving scenario, demonstrating how predictability-aware agents can better coordinate and utilize prediction models. We also observe that agents indirectly exhibit expert-like behaviors, such as following social norms, without explicitly encoding them in the planner. Finally, the fourth experiment explores this direction further by testing interactions with recorded human driver data using a state-of-the-art prediction model, showing that predictability-aware agents achieve safer trajectories as a result of more closely mimicking human behavior.

5.1 Planner

For all experiments in this section, we use a sampling-based planner, namely Model Predictive Path Integral (MPPI) control [29], based on the methodology presented in [36]. MPPI places no restrictions on dynamics model or cost function and converges well toward optima with a moderate amount of samples [35]. Given a nominal control sequence as an initial guess, MPPI applies Gaussian noise at each step to generate a set of M control sequence samples. It then uses a state transition function $f(\cdot)$ to simulate their corresponding M state trajectories. Each of the resulting state trajectories is evaluated based on the cost defined in (5), resulting in a total sample cost J_m. Once J_m, $\forall m \in [1, ..., M]$ is computed, importance sampling weights, w_m, can be calculated as:

$$w_m = \frac{1}{\eta} \exp\left(-\frac{1}{\lambda}(J_m - J_{\min})\right), \quad \sum_{m=1}^{M} w_m = 1,$$

where $J_{\min}$ is the minimum sampled cost, η is a normalization factor and λ is a controlling parameter that controls the width of the weight distribution. These weights prioritize lower-cost trajectories. The optimal control sequence U^* is then calculated as the weighted sum of all sampled control sequences:

$$U^* = \sum_{m=1}^{M} w_m U_m,$$

where we use $U_m = \{\boldsymbol{u}_0, \boldsymbol{u}_1, ..., \boldsymbol{u}_K\}$ to denote the m-th sampled control sequence. It is common to use a time-shifted version of U^* to warm-start the sampling strategy at the next time-step.

5.2 Metrics

As a proxy to measure coordination, we propose the use of planning effort. Planning effort is a metric taken from [7] to quantify how much trajectories deviate from an initial estimate. The authors point out this serves as a proxy for how well the agent is able to anticipate the evolution of its surroundings. We adapt planning effort for receding horizon tasks with the following formulation:

$$\begin{aligned} PE(\xi_{0:T}) &= \frac{1}{T-1} \sum_{t=0}^{T-1} MSE(\tau_t, \tau_{t+1}), \text{ with} \\ MSE(\tau_t, \tau_{t+1}) &= \sum_{k=0}^{K} \|\boldsymbol{x}_k^t - \boldsymbol{x}_k^{t+1}\|, \end{aligned} \tag{6}$$

where T and K denote the total time duration of the simulation and planning horizon length respectively. $\xi_{0:T}$ denotes the set of

all the plans along a trajectory $\xi_{0:T} = \{\tau_0, \tau_1, \ldots, \tau_{T-1}\}$: with τ_t the plan at time-step t. $\boldsymbol{x}_k^t$ represents the state at horizon step k for the plan τ_t. In this context, planning effort measures, on average, the magnitude of an agent's plan update per time-step. Generally, for a given task, a more accurate prediction model corresponds to a lower planning effort.

5.3 Single Agent Experiments

Experiment Objective. In this experiment, we present a single agent interacting with a hand-crafted multi-modal prediction model, serving as a model of an observer's expectation. This is a benchmark task used by previous works on legibility and predictability [14], [26] to provide clear insight into the relationship between predictability and the agent's intrinsic motivation.

Setup. Consider an environment with two possible goals: $\mathcal{G} = \{A : [20, 10], B : [20, -10]\}$. The robot starts at position $\boldsymbol{x}_0 = [0, 0]$ and is tasked with reaching goal B. The predictions model an uncertain observer that holds mistaken initial beliefs $\mathcal{B}$ about the agents goals: $b_0^A = 0.7$ and $b_0^B = 0.3$. Based on these beliefs a Gaussian Mixture $p_t(\boldsymbol{x})$ is used as a prediction, with each mode assuming a Constant Velocity (CV) trajectory towards its respective goal. For timestep t at each horizon step k:

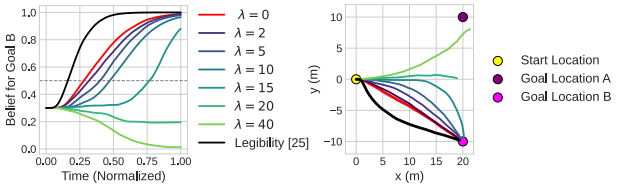
$$p_{t,k}(\boldsymbol{x}) = \sum_{g \in \mathcal{G}} b_t^g p_{t,k}(\boldsymbol{x}), \tag{7}$$

where $p_{t,k}(\boldsymbol{x}) = \mathcal{N}(\mu_t^g, \Sigma)$ with $\mu_{t,k}^g$ is the CV prediction for goal $g \in \mathcal{G}$ at horizon step k, Σ is a fixed covariance and $\boldsymbol{x}$ is a state. We model the observer's changing beliefs $\mathcal{B}$ via Bayesian inference. With every new observation, beliefs are updated using the mode predictions $p_{t,k}(\boldsymbol{x})$ as likelihood functions:

$$b_t^g = \frac{p_{t-1,0}(\boldsymbol{x}_t) b_{t-1}^g}{\sum_{g \in \mathcal{G}} p_{t-1,0}(\boldsymbol{x}_t) b_{t-1}^g}. \tag{8}$$

Keeping a fixed discount $\gamma = 0.6$ in Eq. (5), we vary the magnitude of λ to generate results shown in Figure 1.

Results Discussion. We use this example to study how λ should be tuned to control the trade-off. If predictability dominates (e.g., $\lambda = 20$ or $\lambda = 40$), this results in observations that further reinforce the observer's mistaken belief. It becomes more costly for the robot to pursue its intrinsic motivation with each time-step, thus it fails to complete the task. Conversely, if λ is too low, the robot may still behave unpredictably[1]. For reference, the resulting behavior of an agent optimizing for *legibility* as per the method of [26] is shown as the black line in Figures 1b and 1a. From the perspective of coordination, [26] can be understood as an anticipatory mechanism: By conveying intention in advance, other agents anticipate better in their planning. Our approach similarly mitigates sudden environmental changes, however instead of aiming to directly influence the other agents' beliefs, we rely on a prediction model to avoid the surprising observations throughout the interaction. While this can occasionally result in slightly more costly trajectories for the agent, we achieve similar results without requiring explicit modeling of the other agent, making it more computationally efficient and robust to situations where the agent may not be able to successfully convey its intention. As demonstrated in Figure 1, when the observer's beliefs are misaligned, the agent adopts a pro-social behavior, gently guiding the observer toward the correct belief.



(a) Belief updates over trajectories. **(b) Trajectories for different λ.**

Figure 1: Figure 1a shows that increasing λ effectively decrease the belief update rate for the observer. In Figure 1b, the nominal trajectory is rendered in red. Given the observer holds mistaken initial beliefs about the robot's goal, we observe that increasing the predictability score λ results in trajectories that are more compliant with the observer's expectation.

5.4 Robot-Robot Interactions

5.4.1 *Swapping Tasks.*

Experiment Objective. These experiments explore the benefits of accounting for predictability in robot-robot interactions through swapping-tasks, a common benchmark for robot coordination [3], [38]. By performing tests in an open environment these tests avoid interference of external environmental influences.

Setup. In the experiments, agents are initially positioned on the vertices of a square and tasked with swapping positions with the agent on the opposite vertex (Figure 2a). The optimal solution requires all agents to coordinate by selecting the same collision avoidance strategy, either passing left or right. Additionally, two more scenarios were tested: an asymmetrical swapping task and a double-crossing task, to explore different geometries and interactions. The experiments use a game-theoretic prediction model based on the ALGAMES framework [11], which solves constrained dynamic games to find an optimal joint strategy over a 20-step horizon. The model generates prediction distributions for each horizon step as a Gaussian with user-specified covariance Σ. By testing three values of the predictability parameter λ {0.0, 2.5, 5.0}, we investigate how accounting for predictability impacts agent coordination. Each task was run 50 times, and the results for all three tasks are reported in Table 1 [2]. An illustration comparing the trajectories for all 3 scenarios can be found in Figure 2.

Results Discussion. As seen in Table 1, increasing the predictability parameter λ consistently led to improved performance across all metrics: planning effort (PE), acceleration (Acc), and angular velocity (Ang). Notably, even selecting a small λ causes a pronounced decrease in planning effort, with further increases in λ yielding diminishing returns. This phenomenon can be attributed to the coordination challenge agents face in this environment, which primarily involves equilibrium selection. In situations where agents

[1] In practice, the influence seems to be very dependent on the structure of the main objective cost function, so we recommend tuning λ empirically based on the specific planner and prediction model used.

[2] For $\lambda = 0$ safety constraint violations are low at 1-2 for all the tasks. For higher λ it was 0 for all tasks. As this is not a very informative result it was not included in the tables

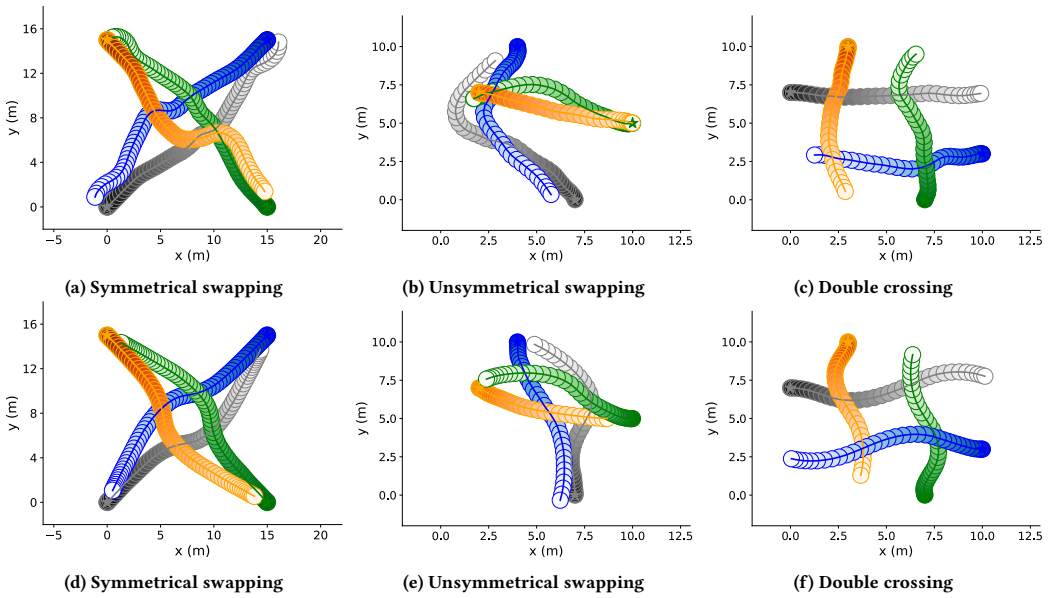


Figure 2: The first row shows the results with $\lambda = 0$ whereas the second row shows the results for $\lambda = 5.0$. When agents account for predictability, aside from faster convergence to a coordination strategy, this also results in smoother trajectories as a consequence of better anticipation of the environment.

Table 1: Table summarizing results for the 3 swapping tasks: Symmetrical, Unsymmetrical, and Double-Crossing

Exp.	Metric	$\lambda = 0.0$	$\lambda = 2.5$	$\lambda = 5.0$
Sym	PE (m²)	2.116 ±1.000	0.516 ±0.161	**0.501 ± 0.145**
	Acc (m/s²)	0.209 ±0.101	**0.038 ± 0.008**	0.043 ±0.009
	Ang (rad/s)	0.283 ±0.041	0.225 ±0.026	**0.219 ± 0.026**
Unsym	PE (m²)	0.877 ±0.489	0.291 ±0.178	**0.187 ± 0.162**
	Acc (m/s²)	0.196 ±0.114	0.138 ±0.090	**0.112 ± 0.080**
	Ang (rad/s)	0.363 ±0.112	0.221 ±0.095	**0.177 ± 0.071**
D-Cross	PE (m²)	0.969 ±0.416	0.388 ±0.124	**0.311 ± 0.116**
	Acc (m/s²)	0.249 ±0.124	**0.123 ± 0.080**	0.125 ±0.070
	Ang (rad/s)	0.434 ±0.128	0.283 ±0.096	**0.252 ± 0.078**

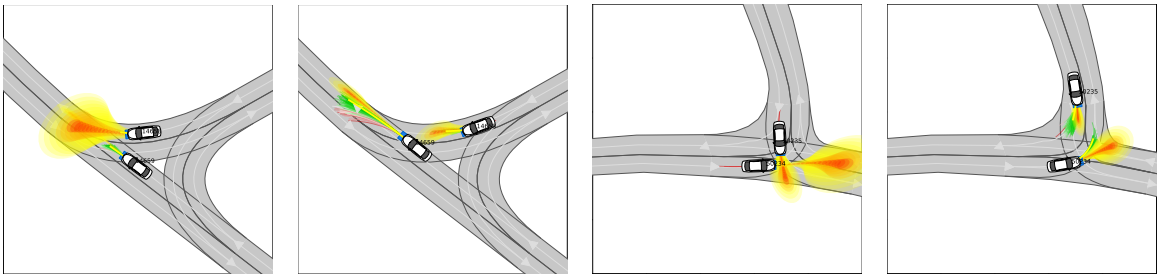
must choose between two equally viable strategies, such as passing left or passing right, our method addresses this challenge by relying on a prediction model to establish a 'soft social convention'. This introduces a subtle bias towards one of the strategies, improving implicit coordination. This mechanism is particularly relevant, as prediction models often excel at capturing an agent's overarching intent and high-level strategy. However, equilibrium selection scenarios are inherently stochastic and unpredictable, making them challenging to model accurately [34]. Thus, our method enhances robustness in such situations by guiding agents towards a coordinated strategy selected by the prediction model. In general, agents need a precise and accurate prediction model for efficient coordination. However, due to the inherent uncertainty of interactions, this is often very hard to achieve. By accounting for predictability, a group of agents is able to establish a 'soft social convention' to mitigate some of this uncertainty. From the perspective of an agent, this results in more accurate predictions, allowing for smoother and more efficient coordination. This mechanism is especially effective for interactions where the main coordination challenge lies in equilibrium selection.

5.4.2 *Robot-Robot Traffic Scenario.*

Experiment Objective. In this experiment, we focus on robot-robot coordination in driving scenarios, where the environment has a stronger influence on agent's behavior. This time, we use a data-driven prediction model to explore how predictability impacts coordination in more complex environments.

Setup. We use CommonRoad [1] as a simulator, which includes the Wale-Net [17] prediction model, a learning-based model that outputs predictions as Gaussians, accounting for uncertainty, road geometry, and the interaction with surrounding agents. Consistent with previous experiments, we employ an MPPI based planner. To account for safety in planning, we implement the constraints introduced by [18], building upon and extending the code from this prior work. We perform tests in two scenarios: A T-Junction and a Lane-Merge. For both scenarios, we perform tests with $\lambda = \{0.0, 2.5, 5.0\}$ for 30 iterations applying small changes in the initial positions and velocities. An illustration of the lane merge environment is presented in Figure 3. The results for T-Junction and lane-merge are presented in Table 2.

Results Discussion. When agents fail to coordinate in road scenarios, they often experience deadlocks or, in the worst case, collisions. In Figure 3a, an example of a deadlock is illustrated. Deadlocks are common in limited space environments such as intersections or narrow passages. Initially, the model may predict one agent will yield while the other advances. However, as deviations occur and both agents hesitate, their predictions begin to reinforce each other's hesitation, creating the deadlock. The model may then be



(a) Lane merge $\lambda = 0$ (b) Lane Merge $\lambda = 5$ (c) T-Junction $\lambda = 0$ (d) T-Junction $\lambda = 5$

Figure 3: a) Illustration of a deadlock With $\lambda = 0$, where a sequence of faulty predictions reinforces both agent's hesitation. b) For $\lambda = 5$, the agents can leverage the prediction model to coordinate which agent gives way and which passes first.

Table 2: Results for T-Junction and Lane Merge Scenarios (Dlk indicates Deadlocks)

Exp	Metric	$\lambda = 0.0$	$\lambda = 2.5$	$\lambda = 5.0$
T-J	Dlk (%)	30.0	**0.0**	**0.0**
	Dist (m)	30.172	**51.248**	47.000
	PE (m^2)	**1.366** ±1.126	2.318 ±0.313	2.507 ±0.759
	Acc (m/s^2)	-0.142 ±0.238	**0.293** ±0.045	0.287 ±0.233
	Ang (rad/s)	0.0037 ±0.0028	**0.0005** ±0.0002	0.0028 ±0.0026
LM	Dlk (%)	73.3	**0.0**	**0.0**
	Dist (m)	46.878	**75.800**	69.909
	PE (m^2)	**2.079** ±0.785	3.513 ±0.564	3.315 ±0.760
	Acc (m/s^2)	**0.111** ±0.092	0.337 ±0.054	0.317 ±0.087
	Ang (rad/s)	0.0032 ±0.0032	0.0009 ±0.0006	**0.0001** ±0.0001

unable to introduce asymmetry to prioritize one of the agents in ambiguous situations, preventing the agents from breaking away from the deadlock. Results show that agents incorporating predictability into their models achieve better coordination, as indicated by less pronounced slowdowns resulting in higher traveled distance and the *disappearance of deadlocks* as seen in Table 2. When examining other metrics, the benefits of incorporating predictability are not as pronounced, especially for higher λ. This occurs because the prediction model is not explicitly conditioned to align with the road geometry (Figures 3c,3d). Since the planner is required to track a reference path, deviations between the predictions and the reference path can push the agent to deviate from the path, requiring small adjustments more frequently for higher higher λ. Although this problem has marginal impact on the overall performance of the agent, the reduction in planning effort may be mitigated.

Similar to the Swapping Task tests, we see that agents are able to use the prediction model to coordinate by reducing uncertainty on equilibrium selection, namely, which agent gives way. However, a noteworthy observation is that, beyond reducing uncertainty, agents enhance their performance by adopting pro-social behaviors embedded in the model's latent space. These behaviors include adherence to social norms and subtle cues learned from training data, mirroring the behavior of experts used to train the model. This behavior resembles imitation learning, where agents learn cooperative strategies directly from expert demonstrations embedded in the prediction model. As seen in Figure 3b, although both outcomes are equally plausible from the raw planning problem, agents consistently converge on the solution where the merging agent yields, which aligns with typical human driving patterns.

5.5 Experiments with human-driver data

Experiment Objective. The goal of this experiment is to evaluate whether predictability can bridge the gap between algorithmic planning and the natural driving patterns observed in humans, facilitating smoother and more adaptive interactions in complex driving environments. We test this by incorporating predictability with simple MPPI-based reference-tracking planner using a SOTA data-driven prediction model.

Setup. We utilize a state-of-the-art (SOTA) prediction model introduced by [23], a multi-modal, transformer-based architecture trained on the Waymo Open Motion Dataset. The model generates scene-centric predictions with three modes, representing the most likely joint trajectories of up to 11 agents, including the ego agent. An MPPI planner is used for reference tracking, incorporating collision avoidance as outlined in [23]. For this experiment, we replay recorded scenes from the Waymo dataset's test set, meaning agents in the environment follow pre-recorded, non-interactive trajectories. The goal is for the ego agent to replicate expert behavior observed during training. We perform tests for $\lambda = 0, 75, 120$[3], over 30 iterations in selected scenarios that require human-like interactions, similar to the approach of [23]. A screenshot of the crossing scenario is shown in Figure 4, and results are reported in Table 3.

Results Discussion. From Table 3, it is evident that increasing the weight of the predictability objective results in fewer collisions and smoother control inputs. Interestingly, however, this does not necessarily correlate with improved progress along the reference path. This can be attributed to the planner inducing less distributional shift in the prediction model. The model, trained on scenes where all agents exhibit expert behaviors, struggles when the planner deviates significantly from these patterns, as it encounters situations outside its training distribution. In such cases, the model attempts to extrapolate and produces sub-optimal predictions, such as incorrectly anticipating that an agent may yield or maneuver differently than it actually does based on the recorded data. This

[3]The large λ values here respond to the particular magnitude of the planning cost function and the prediction model used. We found that values of a higher order of magnitude were needed to obtain predictable behavior shifts.

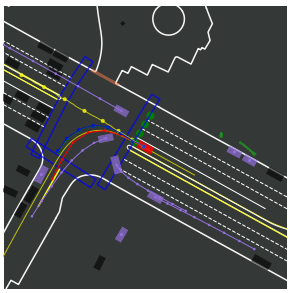


Figure 4: Illustration of the navigation problem in Crossing1. The reference global path is rendered as a smooth yellow line. The AV's plan is rendered in red. Predictions for other agents are rendered in purple, showing only the most likely mode for clarity. The ego-prediction is multi-modal with 3 modes represented by the yellow, green, and blue trajectories.

Table 3: Results comparing the performance of an MPPI-based planner on Waymo Open Motion Dataset scenarios for different λ values. For 30 iterations we present the number of collisions and the mean value of other performance metrics

Scenario	λ	Col (%)	Dist (m)	Acc (m/s²)	Lat_Acc (m/s²)	L2 (m)
Crossing1	0	43.3	74.540 ±1.928	1.085 ±0.121	1.615 ±0.578	4.101 ±0.532
	75	0	68.353 ±0.982	0.681 ±0.025	0.412 ±0.044	3.358 ±0.221
	120	0	55.796 ±1.321	0.472 ±0.035	0.149 ±0.025	2.554 ±0.053
Crossing2	0	86.6	75.843 ±4.803	1.418 ±0.099	2.121 ±0.295	12.707 ±1.112
	75	0	55.353 ±1.403	0.957 ±0.091	0.314 ±0.036	4.603 ±1.246
	120	23.3	37.518 ±12.954	1.534 ±0.736	0.227 ±0.084	2.947 ±3.206
Intersection	0	26.6	69.747 ±4.794	1.450 ±0.153	1.817 ±0.782	24.808 ±3.632
	75	0	72.700 ±0.115	0.709 ±0.029	0.493 ±0.075	24.618 ±1.036
	120	0	71.885 ±0.227	0.613 ±0.043	0.304 ±0.026	19.900 ±0.733
Emergency	0	53.3	61.377 ±20.244	1.258 ±0.175	0.882 ±0.141	8.631 ±5.877
	75	0	68.883 ±0.525	0.763 ±0.010	0.365 ±0.031	1.961 ±0.069
	120	0	60.058 ±0.797	0.581 ±0.019	0.184 ±0.016	1.515 ±0.091

misalignment leads to overconfident behavior in some instances, which, while promoting progress along the reference path, increases the risk of collisions. Evidence supporting this hypothesis is found in the Human L2 loss metric, which measures the L2 loss between the agent's trajectory and the corresponding human trajectory that the planner aims to replicate. For $\lambda = 0$, the higher L2 loss indicates significant deviation from human behavior, suggesting that the agent diverges more from the expert's trajectory. In contrast, when predictability is considered, the L2 loss decreases, indicating that the agent's behavior aligns more closely with the human data. This results in reduced distributional shift and, consequently, more accurate predictions and smoother trajectories.

6 DISCUSSION

Discussion. The method assumes that agents can approximate each other's expectations, often implying a shared prediction model. Although this might seem impractical, certain decentralized settings could accommodate shared models. For instance, in a warehouse environment where multiple Autonomous Ground Vehicles (AGVs) transport valuable goods, a shared prediction model could be feasibly developed and implemented [38]. When integrated with our methodology, such a model could establish 'operational norms', enabling agents to coordinate efficiently and robustly without the need for centralized control, thus reducing computational and infrastructure demands. A comparable scenario is anticipated in future markets where autonomous vehicles (AVs) from different manufacturers must interact. Recent studies pointed at the importance of establishing a unified driving convention [37], as the absence of such a standard could lead to exploitative strategies from different AV companies pursuing competitive advantage, and thereby compromise safety. Different companies can cooperate to develop a shared prediction model to serve as an industry standard. Given a model all AVs in traffic share, our method would enable AVs to anticipate each other's actions and more effectively settle on coordination, thereby providing enhanced road safety without explicit coordination or reliance on infrastructure for centralized coordination.

Future Work. Balancing predictability and performance cost, determined by λ, is complex and context-dependent. Dynamically adjusting λ as agents interact could improve performance, increasingly prioritizing predictability in safety-critical moments. Developing adaptive heuristics for this adjustment, as suggested by previous work [2, 14], would be a valuable research direction. Alternatively, using lexicographic optimization could enhance generalizability by providing a structured trade-off that eliminates the need for tuning a magnitude dependent weighting parameter. However, This requires adpatation of the cost function computation to account for the lexicographic priorities, where predictability is prioritized subject to a performance constraint.

Conclusion. We present a novel approach to enhance multi-agent interaction capabilities for sequential predict-and-plan frameworks by introducing predictability as a key optimization objective. Accounting for predictability in this manner can be understood as an implicit cooperation mechanism whereby agents use a prediction model to actively reduce uncertainty about the environment for other agents. This not only improves the robustness of coordination strategies but also reduces planning effort without requiring explicit communication or high-level control, and does so independently of the number of interacting agents. Through experiments, including robot-robot interactions and human-interaction scenarios, our method improved agent coordination, reduced collisions, and led to smoother, more efficient trajectories (particularly in complex coordination environments). We also demonstrated that the benefits extend to interactions with human drivers by allowing the agent to more reliably use its prediction model.

REFERENCES

[1] Matthias Althoff, Markus Koschi, and Stefanie Manzinger. 2017. CommonRoad: Composable benchmarks for motion planning on roads. In *2017 IEEE Intelligent Vehicles Symposium (IV)*. 719–726. https://doi.org/10.1109/IVS.2017.7995802

[2] Jean-Luc Bastarache, Christopher Nielsen, and Stephen L. Smith. 2023. On Legible and Predictable Robot Navigation in Multi-Agent Environments. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, London, United Kingdom, 5508–5514. https://doi.org/10.1109/ICRA48891.2023.10160572

[3] Maulik Bhatt, Yixuan Jia, and Negar Mehr. 2023. Efficient Constrained Multi-Agent Trajectory Optimization using Dynamic Potential Games. https://doi.org/10.48550/arXiv.2206.08963 arXiv:2206.08963 [cs].

[4] Peter Bossaerts and Carsten Murawski. 2017. Computational Complexity and Human Decision-Making. *Trends in Cognitive Sciences* 21, 12 (Dec. 2017), 917–929. https://doi.org/10.1016/j.tics.2017.09.005 Publisher: Elsevier.

[5] Bruno Brito, Boaz Floor, Laura Ferranti, and Javier Alonso-Mora. 2020. Model Predictive Contouring Control for Collision Avoidance in Unstructured Dynamic Environments. http://arxiv.org/abs/2010.10190 arXiv:2010.10190 [cs].

[6] Tim Brüdigam, Kenan Ahmic, Marion Leibold, and Dirk Wollherr. 2018. Legible Model Predictive Control for Autonomous Driving on Highways. *IFAC-PapersOnLine* 51, 20 (2018), 215–221. https://doi.org/10.1016/j.ifacol.2018.11.016

[7] Daniel Carton, Wiktor Olszowy, and Dirk Wollherr. 2016. Measuring the Effectiveness of Readability for Mobile Robot Locomotion. *International Journal of Social Robotics* 8, 5 (Nov. 2016), 721–741. https://doi.org/10.1007/s12369-016-0358-7

[8] Sergio Casas, Cole Gulino, Simon Suo, Katie Luo, Renjie Liao, and Raquel Urtasun. 2020. Implicit latent variable model for scene-consistent motion forecasting. *European Conference on Computer Vision (ECCV)* (Sept. 2020). https://www.ecva.net/papers/eccv_2020/papers_ECCV/papers/123680613.pdf

[9] Yuxiao Chen, Boris Ivanovic, and Marc Pavone. 2022. Scept: Scene-consistent, policy-based trajectory predictions for planning. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Sept. 2022). https://doi.org/10.1109/CVPR52688.2022.01659

[10] Yuxiao Chen, Sushant Veer, Peter Karkus, and Marco Pavone. 2023. InteractiveMotionPlanning for AutonomousVehicle with Joint Optimization. http://arxiv.org/abs/2310.18301v2

[11] Simon Le Cleac'h, Mac Schwager, and Zachary Manchester. 2020. ALGAMES: A Fast Solver for Constrained Dynamic Games. In *Robotics: Science and Systems XVI*. https://doi.org/10.15607/RSS.2020.XVI.091 arXiv:1910.09713 [cs].

[12] Andrew M. Colman. 2003. Cooperation, psychological game theory, and limitations of rationality in social interaction. *The Behavioral and Brain Sciences* 26, 2 (April 2003), 139–153; discussion 153–198. https://doi.org/10.1017/s0140525x03000050

[13] Oscar de Groot, Laura Ferranti, Dariu Gavrila, and Javier Alonso-Mora. 2023. Scenario-Based Motion Planning with Bounded Probability of Collision. http://arxiv.org/abs/2307.01070 arXiv:2307.01070 [cs].

[14] Anca Dragan and Siddhartha Srinivasa. 2014. Integrating human observer inferences into robot motion planning. *Autonomous Robots* 37, 4 (Dec. 2014), 351–368. https://doi.org/10.1007/s10514-014-9408-x

[15] Anca D. Dragan, Kenton C.T. Lee, and Siddhartha S. Srinivasa. 2013. Legibility and predictability of robot motion. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, Tokyo, Japan, 301–308. https://doi.org/10.1109/HRI.2013.6483603

[16] Jaime F. Fisac, Eli Bronstein, Elis Stefansson, Dorsa Sadigh, S. Shankar Sastry, and Anca D. Dragan. 2018. Hierarchical Game-Theoretic Planning for Autonomous Vehicles. http://arxiv.org/abs/1810.05766 arXiv:1810.05766 [cs, math].

[17] Maximilian Geisslinger, Phillip Karle, Johannes Betz, and Markus Lienkamp. 2021. Watch-and-Learn-Net: Self-supervised Online Learning for Probabilistic Vehicle Trajectory Prediction. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 869–875. https://doi.org/10.1109/SMC52423.2021.9659079 ISSN: 2577-1655.

[18] Maximilian Geisslinger, Franziska Poszler, and Markus Lienkamp. 2023. An ethical trajectory planning algorithm for autonomous vehicles. *Nature Machine Intelligence* 5 (Feb. 2023), 1–8. https://doi.org/10.1038/s42256-022-00607-z

[19] Jasper Geldenbott and Karen Leung. 2024. Legible and Proactive Robot Planning for Prosocial Human-Robot Interactions. http://arxiv.org/abs/2404.03734 arXiv:2404.03734 [cs, eess].

[20] Steffen Hagedorn, Marcel Hallgarten, Martin Stoll, and Alexandru Condurache. 2023. Rethinking Integration of Prediction and Planning in Deep Learning-Based Automated Driving Systems: A Review. http://arxiv.org/abs/2308.05731 arXiv:2308.05731 [cs].

[21] Yanjun Huang, Jiatong Du, Ziru Yang, Zewei Zhou, Lin Zhang, and Hong Chen. 2022. A Survey on Trajectory-Prediction Methods for Autonomous Driving. *IEEE Transactions on Intelligent Vehicles* 7, 3 (Sept. 2022), 652–674. https://doi.org/10.1109/TIV.2022.3167103

[22] Zhiyu Huang, Peter Karkus, Boris Ivanovic, Yuxiao Chen, Marco Pavone, and Chen Lv. 2024. DTPP: DifferentiableJointConditionalPrediction and Cost Evaluation for Tree Policy Planning in Autonomous Driving. http://arxiv.org/abs/2310.05885v2

[23] Zhiyu Huang, Haochen Liu, Jingda Wu, and Chen Lv. 2023. Differentiable Integrated Motion Prediction and Planning With Learnable Cost Function for Autonomous Driving. *IEEE Transactions on Neural Networks and Learning Systems* (2023), 1–15. https://doi.org/10.1109/TNNLS.2023.3283542

[24] David Hyland, Tomáš Gavenčiak, Lancelot Da Costa, Conor Heins, Vojtech Kovarik, Julian Gutierrez, Michael J. Wooldridge, and Jan Kulveit. 2024. Free-Energy Equilibria: Toward a Theory of Interactions Between Boundedly-Rational Agents. In *ICML 2024 Workshop on Models of Human Feedback for AI Alignment*. https://openreview.net/forum?id=4Ft7DcrjdO

[25] Peter Karkus, Boris Ivanovic, Shie Mannor, and Marco Pavone. 2022. DiffStack: A Differentiable and Modular Control Stack for Autonomous Vehicles. http://arxiv.org/abs/2212.06437 arXiv:2212.06437 [cs].

[26] D. Livingston McPherson and S. Shankar Sastry. 2021. An Efficient Understandability Objective for Dynamic Optimal Control. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Prague, Czech Republic, 986–992. https://doi.org/10.1109/IROS51168.2021.9636007

[27] Jiquan Ngiam, Benjamin Caine, Vijay Vasudevan, Zhengdong Zhang, Hao-Tien Lewis Chiang, Jeffrey Ling, Rebecca Roelofs, Alex Bewley, Chenxi Liu, Ashish Venugopal, David Weiss, Ben Sapp, Zhifeng Chen, and Jonathon Shlens. 2022. Scene Transformer: A unified architecture for predicting multiple agent trajectories. http://arxiv.org/abs/2106.08417 arXiv:2106.08417 [cs].

[28] Daniel Jarne Ornia, Giannis Delimpaltadakis, Jens Kober, and Javier Alonso-Mora. 2023. Predictable reinforcement learning dynamics through entropy rate minimization. *arXiv preprint arXiv:2311.18703* (2023).

[29] Corrado Pezzato, Chadi Salmi, Max Spahn, Elia Trevisan, Javier Alonso-Mora, and Carlos Hernandez Corbato. 2023. Sampling-based Model Predictive Control Leveraging Parallelizable Physics Simulations. arXiv:2307.09105 [cs.RO]

[30] Noor Sajid, Philip J Ball, Thomas Parr, and Karl J Friston. 2021. Active inference: demystified and compared. *Neural computation* 33, 3 (2021), 674–712.

[31] Wilko Schwarting, Javier Alonso-Mora, and Daniela Rus. 2018. Planning and Decision-Making for Autonomous Vehicles. *Annual Review of Control, Robotics, and Autonomous Systems* 1, 1 (May 2018), 187–210. https://doi.org/10.1146/annurev-control-060117-105157

[32] Evangelos Theodorou, Jonas Buchli, and Stefan Schaal. 2010. A generalized path integral control approach to reinforcement learning. *The Journal of Machine Learning Research* 11 (2010), 3137–3181.

[33] Elia Trevisan and Javier Alonso-Mora. 2024. Biased-MPPI: Informing Sampling-Based Model Predictive Control by Fusing Ancillary Controllers. *IEEE Robotics and Automation Letters* (2024).

[34] Wenshuo Wang, Letian Wang, Chengyuan Zhang, Changliu Liu, and Lijun Sun. 2022. Social Interactions for Autonomous Driving: A Review and Perspectives. *Foundations and Trends® in Robotics* 10, 3-4 (2022), 198–376. https://doi.org/10.1561/2300000078 arXiv:2208.07541 [cs].

[35] Grady Williams, Andrew Aldrich, and Evangelos A. Theodorou. 2017. Model Predictive Path Integral Control: From Theory to Parallel Computation. *Journal of Guidance, Control, and Dynamics* 40, 2 (Feb. 2017), 344–357. https://doi.org/10.2514/1.G001921

[36] Grady Williams, Paul Drews, Brian Goldfain, James M. Rehg, and Evangelos A. Theodorou. 2017. Information Theoretic Model Predictive Control: Theory and Applications to Autonomous Driving. http://arxiv.org/abs/1707.02342 arXiv:1707.02342 [cs].

[37] Xiaojuan Yu, Vincent A.C. Van Den Berg, Erik T. Verhoef, and Zhi-Chun Li. 2022. Will all autonomous cars cooperate? Brands' strategic interactions under dynamic congestion. *Transportation Research Part E: Logistics and Transportation Review* 166 (Oct. 2022), 102825. https://doi.org/10.1016/j.tre.2022.102825

[38] Hai Zhu, Francisco Martinez Claramunt, Bruno Brito, and Javier Alonso-Mora. 2021. Learning Interaction-Aware Trajectory Predictions for Decentralized Multi-Robot Motion Planning in Dynamic Environments. http://arxiv.org/abs/2102.05382 arXiv:2102.05382 [cs].