STRUCTURED BEHAVIORAL UTILITY LEARNING FOR INTERACTIVE AUTONOMOUS DRIVING

Anonymous authors
Paper under double-blind review

ABSTRACT

Interactive autonomous driving requires forecasts that are not only geometrically accurate but also informative for downstream decision making. Existing prediction–planning pipelines often optimize forecasting in motion space while planning reasons over interaction utility, creating a prediction–utility gap: geometrically similar predictions can still induce different downstream decisions. We propose HSP, a unified framework built on continuous behavioral objective coordinates, which place surrounding agents along structured safety, comfort, efficiency, and social-accommodation axes. Inferred from observed motion, these coordinates condition multimodal prediction, define a trajectory energy for gradient-based refinement, and shape agent-conditioned interaction utilities for ego planning. HSP further introduces a Prospect Value Model, motivated by prospect theory, to learn context-dependent nonlinear mappings from interaction utility to decision value. Experiments across five prediction and planning benchmarks demonstrate strong performance across diverse traffic settings. Under closely matched geometric prediction accuracy, HSP retains clear planning gains, indicating decision-relevant information beyond displacement error. Learned value functions further exhibit context-dependent nonlinear patterns, supporting nonlinear valuation beyond conventional linear utility evaluation.

1 INTRODUCTION

Autonomous driving has made substantial progress in perception, prediction, and control, yet decision making in mixed traffic remains challenging. Interactive driving requires more than forecasting where surrounding agents may move: the ego vehicle must also reason about which aspects of those futures matter for its decisions. Geometric prediction quality and decision relevance, however, are not necessarily aligned. Predictions with similar displacement accuracy may imply different interaction consequences, while small geometric differences can lead to substantially different planning responses in safety-critical or negotiation-intensive scenarios. We refer to this mismatch between motion forecasting and downstream decision relevance as the prediction–utility gap.

Existing behavior-aware prediction methods incorporate maneuver labels, intent categories, driving-style attributes, or latent interaction features to enrich geometric forecasting. However, these representations are typically used as descriptive conditioning variables. Discrete labels provide explicit semantics but limited resolution for continuous behavioral variation, whereas unconstrained latent embeddings are flexible but are not directly tied to how alternative trajectories should be evaluated or optimized. This motivates a behavioral representation that is continuous and semantically structured while remaining operational beyond prediction. Such a representation should not only characterize observed motion, but also support the scoring and refinement of candidate futures and remain directly usable by downstream planning.

A second challenge concerns how interaction utility is translated into decision value. Many planning formulations rank candidate trajectories through additive costs, fixed interaction penalties, or linear utility aggregation (Kolekar et al., 2020). While effective for encoding feasibility and task objectives, these formulations impose relatively rigid valuation of interaction outcomes. Research on bounded rationality and prospect theory suggests that outcome valuation can instead be nonlinear and context dependent (Simon, 1990; Tversky & Kahneman, 1992). Motivated by this perspective,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

we investigate whether interactive planning can benefit from a structured nonlinear valuation mechanism whose shape adapts to the ongoing interaction.

To address these challenges jointly, we propose Human-Social Prospect learning (HSP), a unified framework for structured behavioral utility learning in interactive autonomous driving. HSP represents surrounding agents through continuous behavioral objective coordinates inferred from observed motion along safety, comfort, efficiency, and social-accommodation axes. Unlike behavior labels or prediction-only latent features, these coordinates are operational throughout the pipeline: they condition multimodal trajectory prediction, parameterize a trajectory energy for gradient-based refinement, and shape agent-conditioned interaction utilities for ego planning. The same representation therefore connects behavioral inference, future generation, trajectory optimization, and decision making. HSP further introduces a Prospect Value Model, motivated by prospect theory, that learns context-dependent nonlinear mappings from interaction utility to decision value. Together, these components address the prediction–utility gap at both the representation and valuation levels.

The contributions of this paper are summarized as follows:

- We introduce continuous behavioral objective coordinates, a structured representation that places driving behavior in a continuous objective space. The coordinates combine behavioral semantics with optimization compatibility, allowing the same representation to condition prediction, score candidate futures, and support gradient-based trajectory refinement.
- We develop HSP, a unified prediction–planning framework that propagates behavioral objective coordinates from surrounding-agent forecasting to ego planning, and introduces a prospect value model for context-dependent nonlinear valuation of interaction outcomes.
- We evaluate HSP across multiple prediction and closed-loop planning benchmarks. Beyond overall performance gains, matched-accuracy analysis shows that planning improvements persist when geometric prediction accuracy is controlled, while valuation ablations demonstrate benefits of structured nonlinear decision valuation over linear utility evaluation.

2 Related Work

Behavioral modeling of driving interaction. Driving has long been viewed as a structured behavioral and decision-making process shaped by interaction and adaptation (Michon, 1985; Ranney, 1994; Fuller, 2005; Wang et al., 2022). Accordingly, autonomous-driving methods have incorporated maneuver labels, intent categories, driving-style attributes, and latent interaction embeddings. Explicit labels provide interpretable semantics but discretize behavior into limited categories, whereas latent embeddings offer flexibility but often lack explicit structure (Engström et al., 2024; Hsu et al., 2023). More importantly, these representations are used as predictive conditioning variables rather than operational quantities for evaluating or optimizing alternative trajectories.

Interactive prediction–planning coupling. Trajectory prediction has evolved from recurrent models to graph-based, vectorized, and generative architectures that capture map structure, agent interactions, and multimodal futures (Huang et al., 2021; Girgis et al., 2022; Alahi et al., 2016; Gao et al., 2020; Liao et al., 2024c; Wang et al., 2025a; Liao et al., 2024a; Gu et al., 2022; Wang et al., 2025b). Recent methods increasingly couple prediction with downstream planning or jointly model multi-agent interactions (Sadigh et al., 2016; Huang et al., 2023a; Liao et al., 2025b). However, the interface between forecasting and planning is still commonly represented by predicted trajectories, latent scene features, or task-specific objectives. As a result, geometric forecasting quality does not necessarily reflect the decision relevance of a predicted future.

Interactive decision making. Motion planning in mixed traffic has been studied through optimization, imitation learning, reinforcement learning, and game-theoretic formulations (Trautman & Krause, 2010; Dolgov et al., 2008; Zhang et al., 2020; Scheel et al., 2022; Cheng et al., 2024b; Huang et al., 2023a; Zhang et al., 2025). Interactions are commonly incorporated through handcrafted costs, learned objectives, or socially informed utility terms (Schwarting et al., 2019; Kuderer et al., 2015; Kolekar et al., 2020; Codevilla et al., 2018; Cheng et al., 2024a). When interaction is explicitly represented through utility-based objectives, its contribution is often aggregated through fixed or relatively rigid mappings. Research on bounded rationality and prospect theory instead mo-

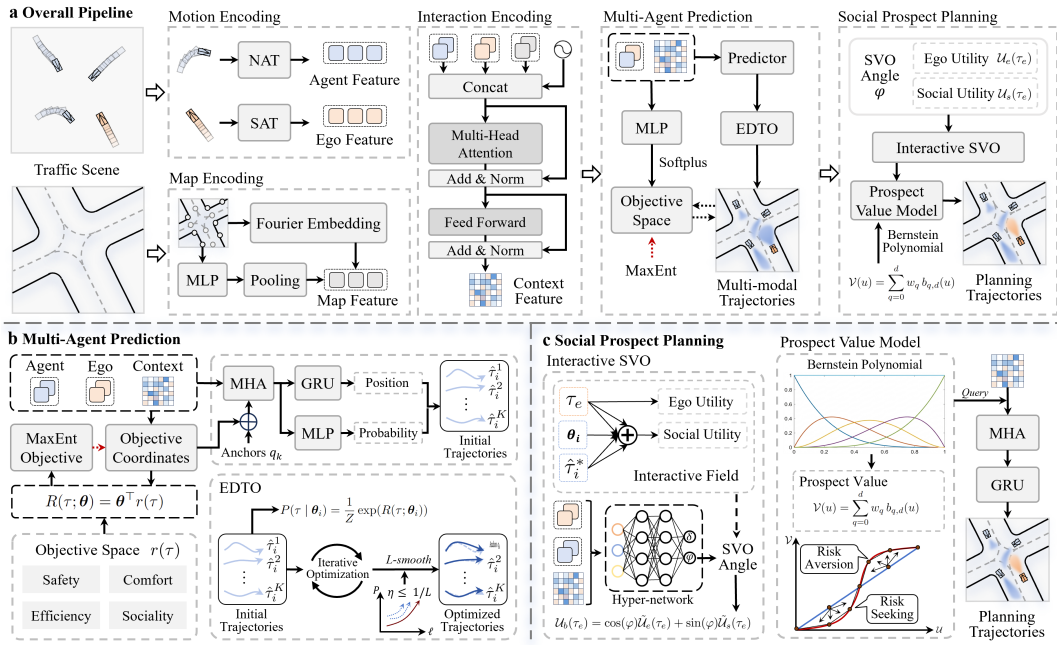

Figure 1: Overview of HSP. (a) Unified prediction–planning pipeline built around Continuous Behavioral Objective Coordinates. (b) Utility-structured multi-agent prediction and trajectory refinement. (c) Social planning with context-adaptive interaction utility and the Prospect Value Model.

tivates nonlinear and context-dependent outcome valuation (Simon, 1990; Tversky & Kahneman, 1992; COLK, 1988), suggesting more flexible interaction valuation beyond fixed linear utility.

3 Methodology

3.1 Problem Formulation and Overview

We study mixed-traffic interaction in a joint prediction–planning setting. At decision step $t = 0$, the system observes the agent histories $\mathcal{H} = \{\mathcal{H}_e, \mathcal{H}_1, \ldots, \mathcal{H}_N\}$ and the high-definition map $\mathcal{M}$, forming the scene context $\mathcal{C} = (\mathcal{H}, \mathcal{M})$. Given $\mathcal{C}$, surrounding-agent prediction estimates K multimodal futures $\hat{\mathcal{T}}_{1:N} = \{\hat{\tau}_i^k\}_{i=1,k=1}^{N,K}$, while ego planning generates K candidate ego trajectories $\hat{\mathcal{T}}_e = \{\hat{\tau}_e^k\}_{k=1}^{K}$ conditioned on both $\mathcal{C}$ and $\hat{\mathcal{T}}_{1:N}$.

As illustrated in Fig. 1, HSP organizes joint prediction and planning around a shared behavioral objective representation. For each surrounding agent, HSP infers continuous behavioral objective coordinates from observed motion over safety, comfort, efficiency, and social-accommodation axes. These coordinates condition multimodal trajectory prediction, parameterize a coordinate-conditioned trajectory energy for gradient-based refinement, and further shape agent-conditioned interaction utilities for ego planning, thereby connecting forecasting, trajectory optimization, and decision making within a shared structured representation. At the planning stage, context-adaptive Social Value Orientation balances ego and social utility, while the Prospect Value Model maps the resulting interaction utility into a context-dependent nonlinear decision value. Together, these components bridge the prediction–utility gap at the representation and valuation levels.

3.2 Scene Encoding

Given the scene context $\mathcal{C} = (\mathcal{H}, \mathcal{M})$, we encode heterogeneous traffic inputs into a unified representation. Surrounding-agent histories are transformed into motion features $F_a = \{F_i\}_{i=1}^{N}$, the ego history is encoded as F_e, and high-definition map elements are encoded as $F_m = \{F_m^j\}_{j=1}^{M}$. The

agent encoders capture temporal motion patterns, while the map encoder represents lane geometry, semantic attributes, and topology. To model global interactions among the ego vehicle, surrounding agents, and map elements, we concatenate all tokens and process them with a multi-layer Transformer encoder ϕ_{TrEnc} using spatial Fourier positional embeddings:

$$F_c = \phi_{TrEnc}\left([F_e \parallel F_a \parallel F_m] + \phi_F(p)\right) \tag{1}$$

where p denotes token coordinates and $\parallel$ denotes concatenation. The final context representation F_c is shared by behavioral coordinate inference, multimodal prediction, and ego planning. Detailed encoder architectures are provided in Appendix A.1.

3.3 Multi-Agent Trajectory Prediction

3.3.1 Continuous Behavioral Objective Coordinates

We represent surrounding-agent behavior in a structured space defined over four interaction-relevant behavioral objectives: safety, comfort, efficiency, and social accommodation. For a candidate trajectory τ, we define its behavioral trajectory score as

$$R(\tau;\boldsymbol{\theta}) = \boldsymbol{\theta}^\top r(\tau), \tag{2}$$

where $\boldsymbol{\theta} \in \mathbb{R}^4_+$ denotes the continuous behavioral objective coordinates and $r(\tau) = [r_{safe}, r_{com}, r_{eff}, r_{soc}]^\top$ collects the trajectory-level objective components. The structured objectives define the basis of the representation, while the coordinates determine how alternative futures are evaluated for each agent, providing a continuous and operational representation of interaction behavior.

Definition 1 (Behavioral Objective Components)**.** *Let $\mathcal{N}$ denote the set of valid neighboring agents. The trajectory-level objective components are defined by:*

$$r_{safe}(\tau) = -\frac{1}{T}\sum_{t=1}^{T}\max_{j\in\mathcal{N}}\left(\frac{[-\Delta v_j^t \cdot n_j]_+}{\|\Delta p_j^t\|_2}\right) \tag{3}$$

$$r_{com}(\tau) = -\frac{1}{T}\sum_{t=1}^{T}\left(\|a^t\|_2^2 + \lambda_{jerk}\|j^t\|_2^2\right) \tag{4}$$

$$r_{eff}(\tau) = \frac{1}{T}\sum_{t=1}^{T}\left(\|v^t\|_2 - \frac{1}{|\mathcal{N}|}\sum_{j\in\mathcal{N}}\|v_j^t\|_2\right) \tag{5}$$

$$r_{soc}(\tau) = -\frac{1}{T}\sum_{t=1}^{T}\frac{1}{|\mathcal{N}|}\sum_{j\in\mathcal{N}}\exp\left(-\frac{1}{2}\left(\frac{(\Delta s_j^t)^2}{(d_0 + h_c\|v^t\|_2)^2} + \frac{(\Delta l_j^t)^2}{w_c^2}\right)\right). \tag{6}$$

Detailed variable definitions for the four behavioral objectives are provided in Appendix A.2.

3.3.2 Objective-Structured Trajectory Prediction

For each surrounding agent i, we infer its behavioral objective coordinates from the contextual motion feature:

$$\boldsymbol{\theta}_i = \phi_{sp}\left(\phi_{coord}(F_i)\right), \tag{7}$$

where ϕ_{coord} is an MLP and $\phi_{sp}(x) = \log(1+e^x)$ enforces non-negative coordinates. The inferred coordinates are integrated with the agent representation, global scene context, and learnable mode anchors to generate multimodal futures structured by the behavioral objective space. Specifically, a cross-attention module first incorporates scene-level interaction information, after which a GRU decoder outputs K candidate trajectories and their mode probabilities:

$$\{(\hat{\tau}_i^k, \hat{\pi}_i^k)\}_{k=1}^K = \phi_{pred}(F_i, F_c, \boldsymbol{\theta}_i). \tag{8}$$

The resulting initial multimodal prediction set is denoted as $\hat{\mathcal{T}}_{init}^i = \{\hat{\tau}_i^1, \ldots, \hat{\tau}_i^K\}$. Detailed decoder implementation is provided in Appendix A.3.

3.3.3 ENERGY-DRIVEN TRAJECTORY OPTIMIZATION

Conditioning the neural decoder on behavioral objective coordinates introduces structured behavioral information into trajectory generation, but the resulting proposals are not explicitly optimized according to the corresponding trajectory score. We therefore use the same coordinates to define an explicit energy over candidate futures, turning the behavioral representation from a predictive conditioning variable into an optimization signal. Following a MaxEnt-inspired formulation, we define

$$P(\tau \mid \boldsymbol{\theta}_i) = \frac{1}{Z(\boldsymbol{\theta}_i)} \exp(R(\tau; \boldsymbol{\theta}_i)), \qquad E(\tau; \boldsymbol{\theta}_i) = -R(\tau; \boldsymbol{\theta}_i). \tag{9}$$

Starting from an initial proposal $\tau^\ell \in \hat{\mathcal{T}}^i_{init}$, we refine the trajectory through gradient-based optimization:

$$\tau^{\ell+1} = \tau^\ell - \eta \nabla_\tau E(\tau^\ell; \boldsymbol{\theta}_i), \tag{10}$$

where η is the step size. The resulting multimodal trajectory set is denoted by $\hat{\mathcal{T}}^i = \{\hat{\tau}^1_i, \ldots, \hat{\tau}^K_i\}$. Collecting the optimized predictions of all surrounding agents yields $\hat{\mathcal{T}}_{1:N}$ for downstream planning. A conditional likelihood monotonicity analysis under a locally smooth energy assumption is provided in Appendix A.4.

3.4 SOCIAL PROSPECT PLANNING

3.4.1 INTERACTIVE SOCIAL VALUE ORIENTATION

Interactive ego planning requires balancing intrinsic driving quality with accommodation to surrounding agents. We therefore construct an ego utility and an agent-conditioned social utility, and combine them through a continuous Social Value Orientation (SVO) angle.

Definition 2 (Ego Utility)**.** *For a candidate ego trajectory τ_e, the ego utility is defined as*

$$\mathcal{U}_e(\tau_e) = \exp\left(\int_0^T \left(\alpha_p \langle \dot{p}_e(t), \vec{\tau}_{ref}(t) \rangle - \alpha_l d^2_{lat}(t) - \alpha_a \|\ddot{p}_e(t)\|^2 - \alpha_j \|\dddot{p}_e(t)\|^2 \right) dt \right). \tag{11}$$

For each neighboring agent, we select its most likely predicted trajectory $\hat{\tau}^*_i = \hat{\tau}_i^{\arg\max_k \hat{\pi}^k_i}$ and construct an agent-conditioned interaction field from its behavioral objective coordinates and predicted future:

$$[A_i, \sigma_i] = \phi_{MLP}\big([\boldsymbol{\theta}_i \parallel \phi_\tau(\hat{\tau}^*_i)]\big), \tag{12}$$

where A_i controls the interaction intensity and σ_i determines its spatial extent.

Definition 3 (Social Utility)**.** *For a candidate ego trajectory τ_e, the social utility is defined as*

$$\mathcal{U}_s(\tau_e) = \exp\left(- \max_{t \in [0,T]} \sum_{i=1}^{N} A_i \exp\left(- \frac{\|p_e(t) - \hat{p}^*_i(t)\|^2}{2\sigma_i^2} \right) \right). \tag{13}$$

After monotone normalization, the two utilities are combined through an SVO angle $\varphi \in [0, \pi/2]$:

$$\mathcal{U}_b(\tau_e) = \cos(\varphi) \tilde{\mathcal{U}}_e(\tau_e) + \sin(\varphi) \tilde{\mathcal{U}}_s(\tau_e), \tag{14}$$

where smaller φ emphasizes ego-driving quality and larger φ places greater weight on social accommodation. We refer to $\mathcal{U}_b$ as the interaction utility. Detailed variable definitions are provided in Appendix A.5.

3.4.2 PROSPECT VALUE MODEL

The interaction utility $\mathcal{U}_b(\tau_e)$ provides a scalar evaluation of an ego trajectory after balancing ego-driving quality and social accommodation. However, equal changes in interaction utility need not translate into equal changes in decision value across different interaction contexts. Motivated by the nonlinear valuation principle of prospect theory, we introduce a Prospect Value Model that transforms normalized interaction utility into decision value.

We first map the interaction utility to a bounded scalar $u = \mathrm{Norm}(\mathcal{U}_b(\tau_e)) \in [0, 1]$, and parameterize the value function using a Bernstein basis:

$$\mathcal{V}(u; \mathbf{w}) = \sum_{q=0}^{d} w_q \, b_{q,d}(u) = \sum_{q=0}^{d} w_q \binom{d}{q} u^q (1-u)^{d-q}, \tag{15}$$

where $b_{q,d}(\cdot)$ are Bernstein basis functions and $\mathbf{w} = [w_0, \ldots, w_d]$ are learnable control points. To impose structural regularity, we set $w_0 = 0$ and construct the remaining control points through cumulative positive increments:

$$w_q = \sum_{j=1}^{q} \delta_j, \qquad \delta_j > 0, \qquad q = 1, \ldots, d. \tag{16}$$

This construction guarantees monotonicity while allowing the curvature of the value function to vary through the increment sequence. The resulting model provides a structured nonlinear mapping from interaction utility to decision value, retaining explicit shape constraints while allowing valuation to depart from a fixed linear transformation. The formal monotonicity and curvature characterization is provided in Appendix A.6.

3.4.3 Context-Adaptive Value Planning

To adapt both social orientation and nonlinear valuation to the ongoing interaction, we encode each neighboring agent using its contextual feature F_i, behavioral objective coordinates $\boldsymbol{\theta}_i$, and selected predicted future $\hat{\tau}_i^*$, and aggregate the resulting interaction tokens with ego-centered attention to obtain F_{int}.

A lightweight hyper-network then generates the context-dependent SVO angle and value-function parameters:

$$[\tilde{\varphi} \parallel \mathbf{z}_\delta] = \phi_{HN}([F_e \parallel F_{int}]), \qquad \varphi = \frac{\pi}{2}\sigma(\tilde{\varphi}), \qquad \delta_q = \ln(1 + e^{(\mathbf{z}_\delta)_q}), \tag{17}$$

with

$$w_0 = 0, \qquad w_q = \sum_{j=1}^{q} \delta_j, \quad q = 1, \ldots, d. \tag{18}$$

The resulting $(\varphi, \mathbf{w})$ jointly determine how ego and social utilities are balanced and how the resulting interaction utility is transformed into decision value. These context-adaptive parameters condition the planner:

$$\{(\hat{\tau}_e^k, \hat{\pi}_e^k)\}_{k=1}^{K} = \phi_{plan}(F_e, F_c, F_{int}, \varphi, \mathbf{w}). \tag{19}$$

Detailed planner architecture is provided in Appendix A.7.

4 Experiments

We evaluate HSP on five benchmarks spanning trajectory prediction and closed-loop planning (Caesar et al., 2020; Deo & Trivedi, 2018; Krajewski et al., 2018; Caesar et al., 2021; Althoff et al., 2017). Dataset protocols, evaluation metrics, training objectives, and implementation details are provided in Appendices B and C.1.

4.1 Prediction and Planning Performance

Trajectory Prediction Results. Table 1 reports the prediction results on nuScenes. HSP achieves the best performance on minADE_5, minFDE_1, minFDE_5, and MR_5, while remaining competitive on minADE_{10}. In particular, HSP reduces minFDE_1 from 6.71 to 6.54 and minFDE_5 from 2.39 to 2.26 relative to the strongest reported baselines on the corresponding metrics. We further evaluate highway prediction on NGSIM and HighD, with full results reported in Appendix C.2.1. HSP obtains the lowest average RMSE on both datasets, improving the strongest competing average from 1.53 to 1.34 on NGSIM and from 0.29 to 0.17 on HighD. The larger gains at longer horizons indicate stronger long-horizon forecasting performance across both urban and highway scenarios.

Table 1: Trajectory prediction performance on nuScenes. **Bold** and underlined text indicate the best and second-best results, respectively. Cases marked with ('-') indicate missing values.

Model	Venue	minADE_5	minADE_{10}	minFDE_1	minFDE_5	MR_5
Trajectron++ (Salzmann et al., 2020)	ECCV 2020	1.88	1.51	9.52	5.63	0.70
LaPred (Kim et al., 2021)	CVPR 2021	1.47	1.12	8.12	3.37	0.53
MHA-JAM (Messaoud et al., 2021)	IV 2021	-	1.18	9.62	3.72	0.64
AutoBot (Girgis et al., 2022)	ICLR 2022	1.37	1.03	8.19	3.40	0.62
THOMAS (Gilles et al., 2022b)	ICLR 2022	1.33	1.04	6.71	-	0.55
PGP (Deo et al., 2022)	CoRL 2022	1.30	1.03	7.17	2.52	0.61
GoHome (Gilles et al., 2022a)	ICRA 2022	1.42	1.15	6.99	-	0.57
EMSIN (Ren et al., 2024)	TFS 2024	1.77	1.28	9.06	3.56	0.54
SeFlow (Zhang et al., 2024b)	ECCV 2024	1.38	**0.98**	7.89	-	0.60
WAKE (Wang et al., 2025a)	TPAMI 2025	1.24	1.09	7.02	2.96	0.55
AMD (Rao et al., 2025)	ICCV 2025	1.23	1.06	6.99	2.43	0.50
NEST (Wang et al., 2025b)	AAAI 2025	1.18	-	6.87	2.39	0.50
HSP (Ours)	-	**1.17**	1.03	**6.54**	**2.26**	**0.49**

Closed-Loop Planning Results. Table 2 reports closed-loop planning results on the *Val14* split of nuPlan. HSP achieves the highest overall scores under both reactive and non-reactive settings, together with strong collision, TTC, and progress scores, showing that the planning improvement maintains both interaction safety and forward progress. HSP also achieves the strongest performance across all reported metrics on CommonRoad, demonstrating consistent cross-benchmark performance under a different scenario distribution and evaluation protocol. Full CommonRoad results and additional nuPlan results on *Test14* and *Test14-hard* are provided in Appendix C.2.2, with computational efficiency and model-complexity analysis reported in Appendix C.3.

Table 2: Closed-loop planning results on the *Val14* split of the nuPlan benchmark. We report the overall closed-loop score under the reactive setting (CLS-R), and the overall score together with detailed sub-metrics under the non-reactive setting (CLS-NR). **Bold** and underlined text indicate the best and second-best results, respectively.

Planner	Venue	CLS-R↑	CLS-NR↑	Collisions↑	TTC↑	Drivable↑	Comfort↑	Progress↑	Speed↑
Log-Replay	-	81.24	93.68	98.76	94.40	98.07	99.27	98.99	96.47
UrbanDriver (Scheel et al., 2022)	CoRL 2021	64.11	68.57	85.60	80.28	90.83	**100.00**	80.83	91.58
PDM-Open (Dauner et al., 2023)	CoRL 2023	54.24	53.53	74.54	69.08	87.89	99.54	69.86	97.72
PDM-Closed (Dauner et al., 2023)	CoRL 2023	92.12	92.84	98.07	93.30	**99.82**	95.52	92.13	**99.83**
GC-PGP (Hallgarten et al., 2023)	ITSC 2023	51.39	61.09	85.87	80.18	89.72	90.00	60.32	99.34
GameFormer(Huang et al., 2023a)	ICCV 2023	79.78	82.95	94.32	86.77	94.87	93.39	89.04	98.67
RasterModel (Karnchanachari et al., 2024)	ICRA 2024	64.37	66.92	86.97	81.46	85.04	81.46	80.60	98.03
IDM (Wong et al., 2024)	CVPR 2024	77.33	79.31	90.92	83.49	94.04	94.40	86.16	97.33
PlanTF (Cheng et al., 2024b)	ICRA 2024	76.95	85.30	94.13	90.73	96.79	93.67	89.83	97.78
CarPlanner(Zhang et al., 2025)	CVPR 2025	87.56	91.45	96.38	91.42	97.73	93.17	91.26	95.48
HSP (Ours)	-	**92.52**	**93.33**	**98.07**	**95.05**	99.63	95.05	**92.69**	98.20

4.2 Ablation Studies

Behavioral Objective Representation. We next examine whether the gains arise from the structured behavioral representation itself rather than additional latent capacity. Table 3 compares the proposed coordinates with an equally dimensional learned latent basis and with shuffled coordinates, where the inferred coordinates are reassigned across agents while the remaining model is unchanged. Replacing the structured representation with a latent basis degrades both prediction and planning, while shuffling the agent–coordinate association causes a substantially larger drop. These results indicate that the benefit depends not only on introducing additional continuous features, but on the structured objective basis and its agent-specific coordinates. Additional analyses of prediction-stage components, dimension-wise objectives, coordinate-space structure, and EDTO refinement and feasibility are provided in Appendix C.4.

Prospect Value Model. We evaluate whether the planning gain arises merely from adding nonlinear function capacity or from the proposed structured valuation. Table 4 compares three variants while keeping the remaining planning pipeline fixed: linear utility, a generic monotonic MLP, and the Prospect Value Model. The monotonic MLP consistently improves over linear valuation, con-

Table 3: Structure-controlled ablation of behavioral objective coordinates. Prediction metrics are evaluated on nuScenes and planning performance is evaluated on nuPlan Test14-hard.

Representation	nuScenes			nuPlan *Test14-hard*
	$\text{minADE}_5 \downarrow$	$\text{minFDE}_5 \downarrow$	$\text{MR}_5 \downarrow$	CLS-R↑
Learned Latent Coordinates	1.21	2.34	0.51	75.87
Shuffled Coordinates	1.24	2.41	0.53	72.96
Behavioral Objective Coordinates	**1.17**	**2.26**	**0.49**	**77.03**

Table 4: Ablation of the decision-value model on nuPlan. We compare linear valuation, a generic monotonic nonlinear mapping, and the proposed Prospect Value Model.

Valuation Function	*Val14*		*Test14*		*Test14-hard*	
	CLS-R ↑	CLS-NR ↑	CLS-R ↑	CLS-NR ↑	CLS-R ↑	CLS-NR ↑
Linear	89.85	90.72	89.34	90.88	71.45	77.35
Monotonic MLP	91.14	92.05	90.62	92.21	73.57	78.94
Prospect Value Model	**92.52**	**93.33**	**91.86**	**93.63**	**77.03**	**82.18**

firming the benefit of relaxing a fixed linear utility mapping. The Prospect Value Model provides a further improvement across all splits, with the largest gain on *Test14-hard*, where CLS-R increases from 71.45 for linear valuation and 73.57 for the monotonic MLP to 77.03. This separation suggests that both nonlinear valuation and its structured, context-adaptive parameterization contribute to planning performance.

Fig. 2 visualizes the learned mapping from normalized interaction utility u to decision value v under different interaction contexts. The learned functions depart systematically from the linear baseline and exhibit context-dependent curvature. In lower-interaction contexts, the mapping shows a pronounced S-shaped profile, whereas higher interaction pressure produces a steeper response in the low-utility region followed by earlier saturation. These patterns are qualitatively consistent with the nonlinear valuation principle motivating prospect theory, while importantly emerging from the learned contextual parameters rather than from a fixed analytic value function. Together with Table 4, the results show that interaction utility benefits from structured and context-dependent nonlinear valuation.

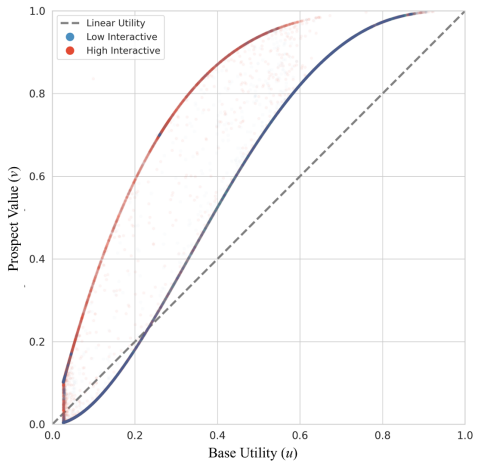

Figure 2: Learned mappings from normalized interaction utility u to decision value v under different interaction contexts.

4.3 Prediction–Utility Analysis

A central motivation of HSP is that geometric forecasting accuracy does not fully characterize the decision relevance of a predicted future. To isolate these two factors, we compare HSP with a geometry-driven predictor–planner baseline on scenarios with approximately matched prediction quality, retaining only cases satisfying $|\Delta\text{ADE}| \leq 0.1$ m. As shown in Fig. 3, the planning-score gain distributions are positively shifted on both *Val14* and *Test14-hard*, with mean gains of 0.20 and 0.39, respectively. These results show that HSP retains downstream planning benefits even when geometric forecasting quality is closely matched, providing direct evidence that displacement accuracy alone does not fully characterize prediction utility. The larger gain on *Test14-hard* further suggests that decision-relevant behavioral information becomes increasingly important in more challenging interaction scenarios. Additional analysis of the relationship between prediction error and planning performance is provided in Appendix C.5.

4.4 Counterfactual Interaction Analysis

To illustrate how the learned behavioral objective coordinates propagate from surrounding-agent motion to downstream ego planning, we conduct a counterfactual intervention on a representative merge scenario, as shown in Fig. 4. In the original case, the key surrounding vehicle accelerates toward the merge region, yielding a coordinate profile with relatively stronger efficiency emphasis and prompting the ego vehicle to adopt a cautious yielding strategy. In the counterfactual case, we modify only the motion of this vehicle so that it decelerates before entering the conflict area, while keeping the scene layout and all other agents unchanged. This intervention shifts the inferred coordinates toward stronger safety emphasis, leading to a different predicted interaction pattern and a corresponding change in ego behavior: the ego no longer needs to wait before the merge, exhibits less deceleration, and follows a smoother trajectory through the interaction region. This example shows that the learned behavioral objective coordinates are not only predictive, but also operational for downstream planning by translating localized changes in surrounding-agent behavior into coherent planning responses. Additional visualizations are provided in Appendix C.6.

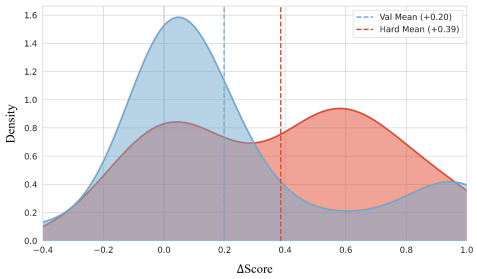

Figure 3: Planning-score gain of HSP over a geometry-driven predictor–planner baseline under matched prediction accuracy ($|\Delta\text{ADE}| \leq 0.1$ m). Both distributions are shifted toward positive gains despite closely matched geometric forecasting accuracy.

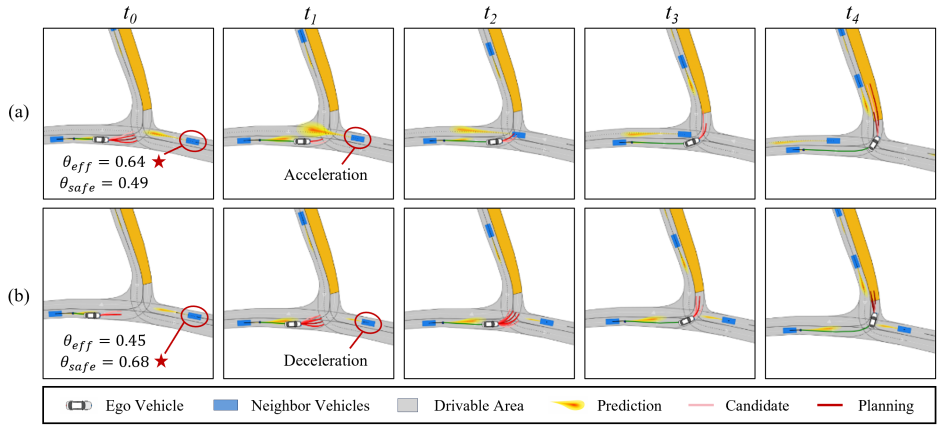

Figure 4: Counterfactual interaction analysis in a merge scenario. (a) Original case. (b) Counterfactual case, where the motion of a key surrounding vehicle is modified from acceleration to deceleration before entering the conflict area. The change in observed motion shifts the inferred behavioral objective coordinates and leads to a different ego planning response.

5 Conclusion

This work introduced HSP, a unified framework for structured behavioral utility learning in interactive autonomous driving. HSP represents surrounding agents through continuous behavioral objective coordinates over safety, comfort, efficiency, and social accommodation, and uses these coordinates to condition multimodal prediction, guide gradient-based trajectory refinement, and inform downstream interaction planning. Experiments across five benchmarks show strong prediction and planning performance, while matched-accuracy analysis indicates that the planning gains extend beyond improvements in geometric forecasting. The Prospect Value Model further shows that context-dependent nonlinear valuation consistently improves over linear and generic monotonic alternatives. Together, these results suggest that connecting motion forecasting and decision making through a structured, optimization-compatible behavioral representation, together with context-adaptive nonlinear interaction valuation, provides an effective formulation for interactive autonomous driving.

AI USE STATEMENT

Generative AI tools were used to assist with early-stage organization and drafting of this manuscript. The authors are responsible for verifying all mathematical derivations, citations, experimental claims, and implementation details, and for revising or removing any AI-assisted content that is not supported by the final study.

REFERENCES

Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. Social lstm: Human trajectory prediction in crowded spaces. pp. 961–971, 2016.

Matthias Althoff, Markus Koschi, and Stefanie Manzinger. Commonroad: Composable benchmarks for motion planning on roads. In *2017 IEEE Intelligent Vehicles Symposium (IV)*, pp. 719–726. IEEE, 2017.

Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *CVPR*, pp. 11621–11631, 2020.

Holger Caesar, Juraj Kabzan, Kok Seang Tan, Whye Kit Fong, Eric Wolff, Alex Lang, Luke Fletcher, Oscar Beijbom, and Sammy Omari. nuplan: A closed-loop ml-based planning benchmark for autonomous vehicles. *arXiv preprint arXiv:2106.11810*, 2021.

Yuning Chai, Benjamin Sapp, Mayank Bansal, and Dragomir Anguelov. Multipath: Multiple probabilistic anchor trajectory hypotheses for behavior prediction. *arXiv preprint arXiv:1910.05449*, 2019.

Xiaobo Chen, Huanjia Zhang, Feng Zhao, Yingfeng Cai, Hai Wang, and Qiaolin Ye. Vehicle trajectory prediction based on intention-aware non-autoregressive transformer with multi-attention learning for internet of vehicles. *IEEE Transactions on Instrumentation and Measurement*, 71: 1–12, 2022a.

Xiaobo Chen, Huanjia Zhang, Feng Zhao, Yu Hu, Chenkai Tan, and Jian Yang. Intention-aware vehicle trajectory prediction based on spatial-temporal dynamic attention network for internet of vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 23(10):19471–19483, 2022b.

Jie Cheng, Yingbing Chen, and Qifeng Chen. Pluto: Pushing the limit of imitation learning-based planning for autonomous driving. *arXiv preprint arXiv:2404.14327*, 2024a.

Jie Cheng, Yingbing Chen, Xiaodong Mei, Bowen Yang, Bo Li, and Ming Liu. Rethinking imitation-based planners for autonomous driving. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 14123–14130. IEEE, 2024b.

Felipe Codevilla, Matthias Müller, Antonio López, Vladlen Koltun, and Alexey Dosovitskiy. End-to-end driving via conditional imitation learning. In *2018 IEEE international conference on robotics and automation (ICRA)*, pp. 4693–4700. IEEE, 2018.

H VAN DER COLK. Risky behaviour resulting from bounded rationality. *Ergonomics*, 31(4): 485–490, 1988.

Daniel Dauner, Marcel Hallgarten, Andreas Geiger, and Kashyap Chitta. Parting with misconceptions about learning-based vehicle motion planning. In *Conference on Robot Learning*, pp. 1268–1281. PMLR, 2023.

Nachiket Deo and Mohan M Trivedi. Multi-modal trajectory prediction of surrounding vehicles with maneuver based lstms. In *2018 IEEE intelligent vehicles symposium (IV)*, pp. 1179–1184. IEEE, 2018.

Nachiket Deo, Eric Wolff, and Oscar Beijbom. Multimodal trajectory prediction conditioned on lane-graph traversals. In *Conference on Robot Learning*, pp. 203–212. PMLR, 2022.

Dmitri Dolgov, Sebastian Thrun, Michael Montemerlo, and James Diebel. Practical search techniques in path planning for autonomous driving. *ann arbor*, 1001(48105):18–80, 2008.

Johan Engström, Ran Wei, Anthony D McDonald, Alfredo Garcia, Matthew O'Kelly, and Leif Johnson. Resolving uncertainty on the fly: modeling adaptive driving behavior as active inference. *Frontiers in neurorobotics*, 18:1341750, 2024.

Ray Fuller. Towards a general theory of driver behaviour. *Accident analysis & prevention*, 37(3): 461–472, 2005.

Jiyang Gao, Chen Sun, Hang Zhao, Yi Shen, Dragomir Anguelov, Congcong Li, and Cordelia Schmid. Vectornet: Encoding hd maps and agent dynamics from vectorized representation. pp. 11525–11533, 2020.

Kai Gao, Xunhao Li, Bin Chen, Lin Hu, Jian Liu, Ronghua Du, and Yongfu Li. Dual transformer based prediction for lane change intentions and trajectories in mixed traffic environment. *IEEE Transactions on Intelligent Transportation Systems*, 24(6):6203–6216, 2023.

Thomas Gilles, Stefano Sabatini, Dzmitry Tsishkou, Bogdan Stanciulescu, and Fabien Moutarde. Gohome: Graph-oriented heatmap output for future motion estimation. In *2022 international conference on robotics and automation (ICRA)*, pp. 9107–9114. IEEE, 2022a.

Thomas Gilles, Stefano Sabatini, Dzmitry Tsishkou, Bogdan Stanciulescu, and Fabien Moutarde. Thomas: Trajectory heatmap output with learned multi-agent sampling. In *International Conference on Learning Representations*, 2022b.

Roger Girgis, Florian Golemo, Felipe Codevilla, Martin Weiss, Jim Aldon D'Souza, Samira Ebrahimi Kahou, Felix Heide, and Christopher Pal. Latent variable sequential set transformers for joint multi-agent motion prediction. In *International Conference on Learning Representations*, 2022.

Tianpei Gu, Guangyi Chen, Junlong Li, Chunze Lin, Yongming Rao, Jie Zhou, and Jiwen Lu. Stochastic trajectory prediction via motion indeterminacy diffusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 17113–17122, 2022.

Marcel Hallgarten, Martin Stoll, and Andreas Zell. From prediction to planning with goal conditioned lane graph traversals. In *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 951–958. IEEE, 2023.

Ali Hassani, Steven Walton, Jiachen Li, Shen Li, and Humphrey Shi. Neighborhood attention transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 6185–6194, 2023.

Kai-Chieh Hsu, Karen Leung, Yuxiao Chen, Jaime F Fisac, and Marco Pavone. Interpretable trajectory prediction for autonomous vehicles via counterfactual responsibility. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5918–5925. IEEE, 2023.

Mengyang Huang, Menggang Zhu, Yunpeng Xiao, and Yanbing Liu. Bayonet-corpus: a trajectory prediction method based on bayonet context and bidirectional gru. *Digital Communications and Networks*, 7(1):72–81, 2021.

Zhiyu Huang, Haochen Liu, and Chen Lv. Gameformer: Game-theoretic modeling and learning of transformer-based interactive prediction and planning for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3903–3913, 2023a.

Zhiyu Huang, Haochen Liu, Jingda Wu, and Chen Lv. Differentiable integrated motion prediction and planning with learnable cost function for autonomous driving. *IEEE transactions on neural networks and learning systems*, 35(11):15222–15236, 2023b.

Napat Karnchanachari, Dimitris Geromichalos, Kok Seang Tan, Nanxiang Li, Christopher Eriksen, Shakiba Yaghoubi, Noushin Mehdipour, Gianmarco Bernasconi, Whye Kit Fong, Yiluan Guo, et al. Towards learning-based planning: The nuplan benchmark for real-world autonomous driving. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 629–636. IEEE, 2024.

ByeoungDo Kim, Seong Hyeon Park, Seokhwan Lee, Elbek Khoshimjonov, Dongsuk Kum, Junsoo Kim, Jeong Soo Kim, and Jun Won Choi. Lapred: Lane-aware prediction of multi-modal future trajectories of dynamic agents. pp. 14636–14645, 2021.

Sarvesh Kolekar, Joost De Winter, and David Abbink. Human-like driving behaviour emerges from a risk-based driver model. *Nature communications*, 11(1):4850, 2020.

Robert Krajewski, Julian Bock, Laurent Kloeker, and Lutz Eckstein. The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems. In *2018 21st international conference on intelligent transportation systems (ITSC)*, pp. 2118–2125. IEEE, 2018.

Markus Kuderer, Shilpa Gulati, and Wolfram Burgard. Learning driving styles for autonomous vehicles from demonstration. In *2015 IEEE international conference on robotics and automation (ICRA)*, pp. 2641–2646. IEEE, 2015.

Haicheng Liao, Zhenning Li, Huanming Shen, Wenxuan Zeng, Dongping Liao, Guofa Li, and Chengzhong Xu. Bat: Behavior-aware human-like trajectory prediction for autonomous driving. volume 38, pp. 10332–10340, 2024a.

Haicheng Liao, Shangqian Liu, Yongkang Li, Zhenning Li, Chengyue Wang, Yunjian Li, Shengbo Eben Li, and Chengzhong Xu. Human observation-inspired trajectory prediction for autonomous driving in mixed-autonomy traffic environments. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 14212–14219. IEEE, 2024b.

Haicheng Liao, Chengyue Wang, Zhenning Li, Yongkang Li, Bonan Wang, Guofa Li, and Chengzhong Xu. Physics-informed trajectory prediction for autonomous driving under missing observation. In *IJCAI*, 2024c.

Haicheng Liao, Zhenning Li, Guohui Zhang, Keqiang Li, and Chengzhong Xu. Toward human-like trajectory prediction for autonomous driving: A behavior-centric approach. *Transportation Science*, 2025a.

Haicheng Liao, Zhenning Li, Kaiqun Zhu, Keqiang Li, and Chengzhong Xu. Sa-tpˆ{2}: A safety-aware trajectory prediction and planning model for autonomous driving. *IEEE Transactions on Robotics*, 2025b.

Mengmeng Liu, Hao Cheng, Lin Chen, Hellward Broszio, Jiangtao Li, Runjiang Zhao, Monika Sester, and Michael Ying Yang. Laformer: Trajectory prediction for autonomous driving with lane-aware scene constraints. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2039–2049, 2024.

Kaouther Messaoud, Nachiket Deo, Mohan M Trivedi, and Fawzi Nashashibi. Trajectory prediction for autonomous driving based on multi-head attention with joint agent-map representation. In *2021 IEEE intelligent vehicles symposium (IV)*, pp. 165–170. IEEE, 2021.

John A Michon. A critical view of driver behavior models: what do we know, what should we do? In *Human behavior and traffic safety*, pp. 485–524. Springer, 1985.

Seokha Moon, Hyun Woo, Hongbeen Park, Haeji Jung, Reza Mahjourian, Hyung-gun Chi, Hyerin Lim, Sangpil Kim, and Jinkyu Kim. Visiontrap: Vision-augmented trajectory prediction guided by textual descriptions. In *European Conference on Computer Vision*, pp. 361–379. Springer, 2024.

Thomas A Ranney. Models of driving behavior: a review of their evolution. *Accident analysis & prevention*, 26(6):733–750, 1994.

Bin Rao, Haicheng Liao, Yanchen Guan, Chengyue Wang, Bonan Wang, Jiaxun Zhang, and Zhenning Li. Amd: Adaptive momentum and decoupled contrastive learning framework for robust long-tail trajectory prediction. *arXiv preprint arXiv:2507.01801*, 2025.

Yilong Ren, Zhengxing Lan, Lingshan Liu, and Haiyang Yu. Emsin: Enhanced multistream interaction network for vehicle trajectory prediction. *IEEE Transactions on Fuzzy Systems*, 33(1):54–68, 2024.

Dorsa Sadigh, Shankar Sastry, Sanjit A Seshia, and Anca D Dragan. Planning for autonomous cars that leverage effects on human actions. In *Robotics: Science and systems*, volume 2, pp. 1–9. Ann Arbor, MI, USA, 2016.

Tim Salzmann, Boris Ivanovic, Punarjay Chakravarty, and Marco Pavone. Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data. In *European Conference on Computer Vision*, pp. 683–700. Springer, 2020.

Oliver Scheel, Luca Bergamini, Maciej Wolczyk, Błażej Osiński, and Peter Ondruska. Urban driver: Learning to drive from real-world demonstrations using policy gradients. In *Conference on Robot Learning*, pp. 718–728. PMLR, 2022.

Wilko Schwarting, Alyssa Pierson, Javier Alonso-Mora, Sertac Karaman, and Daniela Rus. Social behavior for autonomous vehicles. *Proceedings of the National Academy of Sciences*, 116(50): 24972–24978, 2019.

Herbert A Simon. Bounded rationality. In *Utility and probability*, pp. 15–18. Springer, 1990.

Peter Trautman and Andreas Krause. Unfreezing the robot: Navigation in dense, interacting crowds. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 797–803. IEEE, 2010.

Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4):297–323, 1992.

Chengyue Wang, Haicheng Liao, Zhenning Li, and Chengzhong Xu. Wake: Towards robust and physically feasible trajectory prediction for autonomous vehicles with wavelet and kinematics synergy. *PAMI*, 2025a.

Chengyue Wang, Haicheng Liao, Bonan Wang, Yanchen Guan, Bin Rao, Ziyuan Pu, Zhiyong Cui, Cheng-Zhong Xu, and Zhenning Li. Nest: A neuromodulated small-world hypergraph trajectory prediction model for autonomous driving. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 808–816, 2025b.

Chengyue Wang, Haicheng Liao, Kaiqun Zhu, Guohui Zhang, and Zhenning Li. A dynamics-enhanced learning model for multi-horizon trajectory prediction in autonomous vehicles. *Information Fusion*, 118:102924, 2025c.

Mingqiang Wang, Lei Zhang, Jun Chen, Zhiqiang Zhang, Zhenpo Wang, and Dongpu Cao. A hybrid trajectory prediction framework for automated vehicles with attention mechanisms. *IEEE Transactions on Transportation Electrification*, 10(3):6178–6194, 2023a.

Renzhi Wang, Senzhang Wang, Hao Yan, and Xiang Wang. Wsip: Wave superposition inspired pooling for dynamic interactions-aware trajectory prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 4685–4692, 2023b.

Wenshuo Wang, Letian Wang, Chengyuan Zhang, Changliu Liu, and Lijun Sun. Social interactions for autonomous driving: A review and perspectives. *Foundations and Trends® in Robotics*, 10 (3-4):198–377, 2022.

Conghao Wong, Beihao Xia, Ziqian Zou, Yulong Wang, and Xinge You. Socialcircle: Learning the angle-based social interaction representation for pedestrian trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19005–19015, 2024.

Xu Xie, Chi Zhang, Yixin Zhu, Ying Nian Wu, and Song-Chun Zhu. Congestion-aware multi-agent trajectory prediction for collision avoidance. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 13693–13700. IEEE, 2021.

Ye Yuan, Xinshuo Weng, Yanglan Ou, and Kris M Kitani. Agentformer: Agent-aware transformers for socio-temporal multi-agent forecasting. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9813–9823, 2021.

Dongkun Zhang, Jiaming Liang, Sha Lu, Ke Guo, Qi Wang, Rong Xiong, Zhenwei Miao, and Yue Wang. Pep: Policy-embedded trajectory planning for autonomous driving. *IEEE Robotics and Automation Letters*, 9(12):11361–11368, 2024a.

Dongkun Zhang, Jiaming Liang, Ke Guo, Sha Lu, Qi Wang, Rong Xiong, Zhenwei Miao, and Yue Wang. Carplanner: Consistent auto-regressive trajectory planning for large-scale reinforcement learning in autonomous driving. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 17239–17248, 2025.

Qingwen Zhang, Yi Yang, Peizheng Li, Olov Andersson, and Patric Jensfelt. Seflow: A self-supervised scene flow method in autonomous driving. In *European Conference on Computer Vision*, pp. 353–369. Springer, 2024b.

Xiaojing Zhang, Alexander Liniger, and Francesco Borrelli. Optimization-based collision avoidance. *IEEE Transactions on Control Systems Technology*, 29(3):972–983, 2020.

Zikang Zhou, Luyao Ye, Jianping Wang, Kui Wu, and Kejie Lu. Hivt: Hierarchical vector transformer for multi-agent motion prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8823–8833, 2022.

Appendix

A METHOD DETAILS

A.1 SCENE ENCODING

To obtain a unified representation of the traffic scene, we construct a hierarchical encoder that integrates agent motion encoding, map encoding, and global interaction encoding. Given the observed motion histories and high-definition map elements, this encoder transforms heterogeneous scene inputs into a shared contextual representation that supports both downstream trajectory prediction and ego-vehicle planning.

A.1.1 AGENT MOTION ENCODING

For each surrounding agent, the observed motion history is first converted into step-wise temporal differences to explicitly highlight local kinematic variations. To encode temporal dependencies, we employ a Neighborhood Attention Transformer (NAT) (Hassani et al., 2023) as the sequence encoder ϕ_T. Unlike global self-attention, NAT restricts attention to a sliding local temporal window, which is well suited to modeling short-term motion continuity while reducing the influence of temporally distant noise. In addition, the discrete semantic category c of each agent is mapped through a learnable embedding layer ϕ_E. The resulting motion feature of a surrounding agent is written as:

$$F_i = \phi_T(\Delta\mathcal{H}_i) + \phi_E(c_i) \tag{20}$$

where the encoded features of all surrounding agents are aggregated as $F_a = \{F_i\}_{i=1}^{N}$.

For the ego vehicle, we retain the absolute historical state sequence rather than temporal differences, since its full state evolution is directly relevant to downstream planning. We encode the ego history $\mathcal{H}_e$ using a dedicated multi-head self-attention module ϕ_{SA} under stochastic feature masking. The resulting ego feature is:

$$F_e = \phi_{SA}(\mathcal{H}_e) + \phi_E(c_e) \tag{21}$$

A.1.2 MAP ENCODING

Within the high-definition map, each structural element is represented as a set of polygonal vectors. To obtain compact geometric descriptors, we employ a permutation-invariant encoder ϕ_P, which processes the local point set of each map element through a multilayer perceptron (MLP) followed by max pooling. Its discrete semantic attribute s is embedded by ϕ_E, and its continuous topological attribute r is encoded by a Fourier embedding function ϕ_F. The resulting feature of a map element is:

$$F_m^j = \phi_P(\mathcal{M}_j) + \phi_E(s_j) + \phi_F(r_j) \tag{22}$$

where the encoded features of all map elements are aggregated as $F_m = \{F_m^j\}_{j=1}^{M}$.

A.1.3 INTERACTION ENCODING

To model global interactions among surrounding agents, the ego vehicle, and static map elements, we concatenate the ego feature F_e, the surrounding-agent features F_a, and the map features F_m into a unified token sequence. To preserve spatial structure, we inject continuous Fourier positional encodings generated from the global coordinates of all tokens. The resulting sequence is then processed by a multi-layer Transformer encoder ϕ_{TrEnc} to obtain the final scene context representation:

$$F_c = \phi_{TrEnc}\left([F_e \parallel F_a \parallel F_m] + \phi_F(p)\right) \tag{23}$$

where $\parallel$ denotes sequence concatenation and p denotes the global coordinates associated with each token. The encoded feature F_c serves as the shared scene context for subsequent behavioral coordinate inference modeling, surrounding-agent prediction, and ego-vehicle planning.

A.2 Continuous Behavioral Objective Coordinates

We define a structured representation over four interaction-relevant behavioral objectives: safety, comfort, efficiency, and social accommodation. For a candidate trajectory τ, its behavioral trajectory score is defined as

$$R(\tau;\boldsymbol{\theta}) = \boldsymbol{\theta}^\top r(\tau), \tag{24}$$

where $\boldsymbol{\theta} \in \mathbb{R}^4_+$ denotes the continuous behavioral objective coordinates and $r(\tau) = [r_{safe}, r_{com}, r_{eff}, r_{soc}]^\top$ collects the trajectory-level objective components. The four objectives define a structured behavioral basis, whereas the agent-specific coordinates are inferred from observed motion and determine their relative contributions when evaluating alternative futures.

Let $\mathcal{N}$ denote the set of valid neighboring agents. The four behavioral objective components are defined as follows.

Definition (Safety Objective Component)**.** *The safety component penalizes trajectories involving stronger closing interactions at small spatial clearance:*

$$r_{safe}(\tau) = -\frac{1}{T}\sum_{t=1}^{T}\max_{j\in\mathcal{N}}\left(\frac{[-\Delta v_j^t \cdot n_j]_+}{\|\Delta p_j^t\|_2}\right), \tag{25}$$

where Δp_j^t and Δv_j^t are the relative position and velocity with respect to neighbor j, n_j is the unit direction toward j, and $[x]_+ = \max(0, x)$ extracts closing interactions.

Definition (Comfort Objective Component)**.** *The comfort component favors dynamically smooth trajectories by penalizing acceleration and jerk:*

$$r_{com}(\tau) = -\frac{1}{T}\sum_{t=1}^{T}\left(\|a^t\|_2^2 + \lambda_{jerk}\|j^t\|_2^2\right), \tag{26}$$

where a^t and j^t denote acceleration and jerk, respectively.

Definition (Efficiency Objective Component)**.** *The efficiency component measures trajectory progress relative to surrounding traffic:*

$$r_{eff}(\tau) = \frac{1}{T}\sum_{t=1}^{T}\left(\|v^t\|_2 - \frac{1}{|\mathcal{N}|}\sum_{j\in\mathcal{N}}\|v_j^t\|_2\right), \tag{27}$$

where v^t is the target-agent velocity and v_j^t is the velocity of neighbor j.

Definition (Social-Accommodation Objective Component)**.** *The social-accommodation component penalizes trajectories that intrude into the interaction space of surrounding traffic:*

$$r_{soc}(\tau) = -\frac{1}{T}\sum_{t=1}^{T}\frac{1}{|\mathcal{N}|}\sum_{j\in\mathcal{N}}\exp\left(-\frac{1}{2}\left(\frac{(\Delta s_j^t)^2}{(d_0 + h_c\|v^t\|_2)^2} + \frac{(\Delta l_j^t)^2}{w_c^2}\right)\right), \tag{28}$$

where Δs_j^t and Δl_j^t are the longitudinal and lateral relative positions in the local driving frame, and d_0, h_c, and w_c control the base distance, speed-adaptive headway, and lateral interaction width.

A.3 Objective-Structured Trajectory Prediction

For each surrounding agent i with contextual representation F_i, we infer its behavioral objective coordinates from the observed motion context:

$$\boldsymbol{\theta}_i = \phi_{\mathrm{sp}}\left(\phi_{\mathrm{coord}}(F_i)\right), \tag{29}$$

where ϕ_{coord} denotes an MLP and $\phi_{\mathrm{sp}}(x) = \log(1 + e^x)$ is the Softplus function, which enforces non-negative coordinates. The resulting $\boldsymbol{\theta}_i$ specifies the agent's continuous coordinates along the safety, comfort, efficiency, and social-accommodation objectives.

Given $\boldsymbol{\theta}_i$, we construct an objective-structured decoder for multimodal trajectory prediction. The target agent first queries the global scene context F_c through a multi-head cross-attention module:

$$h_i = \phi_{\mathrm{LN}}\left(F_i + \phi_{\mathrm{Attn}}(F_i, F_c, F_c)\right), \tag{30}$$

where ϕ_{LN} denotes Layer Normalization. For each prediction mode $k \in \{1, \ldots, K\}$, we fuse the interaction-aware context h_i, learnable mode anchor q_k, and behavioral objective coordinates $\boldsymbol{\theta}_i$ into a unified latent state:

$$z_i^k = \phi_{\mathrm{fuse}}\left([h_i \parallel q_k \parallel \boldsymbol{\theta}_i]\right), \tag{31}$$

where $\parallel$ denotes concatenation.

The latent state initializes a Gated Recurrent Unit (GRU) decoder ϕ_{GRU}. At each future step t, the decoder updates its hidden state $s_i^{k,t}$ from the previous state and decoder input $u_i^{k,t}$:

$$s_i^{k,t} = \phi_{\mathrm{GRU}}(u_i^{k,t}, s_i^{k,t-1}), \qquad s_i^{k,0} = z_i^k. \tag{32}$$

Here, $u_i^{k,t}$ is obtained from an embedding of the previously predicted state.

A trajectory regression head ϕ_{loc} predicts the future position $\hat{p}_i^{k,t}$, while a classification head ϕ_{cls} estimates the probability of each prediction mode:

$$\hat{p}_i^{k,t} = \phi_{\mathrm{loc}}(s_i^{k,t}), \qquad \hat{\pi}_i^k = \frac{\exp(\phi_{\mathrm{cls}}(z_i^k))}{\sum_{k'=1}^{K} \exp(\phi_{\mathrm{cls}}(z_i^{k'}))}. \tag{33}$$

The resulting initial multimodal trajectory set is

$$\hat{\mathcal{T}}_{\mathrm{init}}^i = \{\hat{\tau}_i^1, \ldots, \hat{\tau}_i^K\}, \tag{34}$$

where $\hat{\tau}_i^k = \{\hat{p}_i^{k,t}\}_{t=1}^T$ denotes the predicted future trajectory of agent i under mode k.

A.4 EDTO Likelihood Monotonicity

Theorem 1 (EDTO Likelihood Monotonicity)**.** *Fix the behavioral objective coordinates $\boldsymbol{\theta}_i$ and let $E(\tau) = E(\tau; \boldsymbol{\theta}_i)$. Assume that E is L-smooth on a neighborhood containing the segment between τ^ℓ and $\tau^{\ell+1}$. Under the update in Eq. 10, if $0 < \eta \leq 1/L$, then the coordinate-conditioned trajectory likelihood is non-decreasing:*

$$P(\tau^{\ell+1} \mid \boldsymbol{\theta}_i) \geq P(\tau^\ell \mid \boldsymbol{\theta}_i). \tag{35}$$

Moreover, if $\|\nabla_\tau E(\tau^\ell)\|_2 > 0$, then

$$P(\tau^{\ell+1} \mid \boldsymbol{\theta}_i) \geq P(\tau^\ell \mid \boldsymbol{\theta}_i) \exp\left(\frac{\eta}{2}\|\nabla_\tau E(\tau^\ell)\|_2^2\right). \tag{36}$$

Proof. Since $E(\tau)$ is L-smooth on a neighborhood containing the segment between τ^ℓ and $\tau^{\ell+1}$, the descent lemma gives

$$E(\tau^{\ell+1}) \leq E(\tau^\ell) + \langle \nabla_\tau E(\tau^\ell), \tau^{\ell+1} - \tau^\ell \rangle + \frac{L}{2}\|\tau^{\ell+1} - \tau^\ell\|_2^2. \tag{37}$$

Substituting Eq. 10 yields

$$E(\tau^{\ell+1}) \leq E(\tau^\ell) - \eta\left(1 - \frac{L\eta}{2}\right)\|\nabla_\tau E(\tau^\ell)\|_2^2. \tag{38}$$

Since $\eta \leq 1/L$,

$$1 - \frac{L\eta}{2} \geq \frac{1}{2}, \tag{39}$$

and therefore

$$E(\tau^{\ell+1}) - E(\tau^\ell) \leq -\frac{\eta}{2}\|\nabla_\tau E(\tau^\ell)\|_2^2. \tag{40}$$

Under the MaxEnt formulation,

$$\log P(\tau \mid \boldsymbol{\theta}_i) = -E(\tau; \boldsymbol{\theta}_i) - \log Z(\boldsymbol{\theta}_i), \tag{41}$$

where $Z(\boldsymbol{\theta}_i)$ is independent of τ. Hence,

$$\begin{aligned} &\log P(\tau^{\ell+1} \mid \boldsymbol{\theta}_i) - \log P(\tau^\ell \mid \boldsymbol{\theta}_i) \\ &\quad = -\big(E(\tau^{\ell+1}) - E(\tau^\ell)\big) \\ &\quad \geq \frac{\eta}{2}\|\nabla_\tau E(\tau^\ell)\|_2^2. \end{aligned} \tag{42}$$

It follows that $P(\tau^{\ell+1} \mid \boldsymbol{\theta}_i) \geq P(\tau^\ell \mid \boldsymbol{\theta}_i)$. When $\|\nabla_\tau E(\tau^\ell)\|_2 > 0$, exponentiating the inequality yields Eq. 36. □

A.5 Detailed Definitions of SVO Utilities

In interactive traffic, ego planning balances intrinsic driving quality with accommodation to surrounding agents. We therefore define an ego utility and an agent-conditioned social utility, and combine them through a continuous Social Value Orientation (SVO) angle.

Definition (Ego Utility)**.** *For a candidate ego trajectory τ_e, the ego utility measures route progress, route adherence, and kinematic smoothness:*

$$\mathcal{U}_e(\tau_e) = \exp\left(\int_0^T \left(\alpha_p \langle \dot{p}_e(t), \vec{\tau}_{ref}(t)\rangle - \alpha_l d_{lat}^2(t) - \alpha_a \|\ddot{p}_e(t)\|^2 - \alpha_j \|\dddot{p}_e(t)\|^2\right) dt\right). \quad (43)$$

Here, $\langle \dot{p}_e(t), \vec{\tau}_{ref}(t)\rangle$ denotes progress along the reference route, $d_{lat}(t)$ is the lateral deviation from the route centerline, and the acceleration and jerk terms regularize motion smoothness.

For each neighboring agent i, we select its most likely predicted trajectory $\hat{\tau}_i^* = \hat{\tau}_i^{\arg\max_k \hat{\pi}_i^k}$. Its behavioral objective coordinates $\boldsymbol{\theta}_i$ and predicted future $\hat{\tau}_i^*$ are then used to construct an agent-conditioned interaction field:

$$[A_i, \sigma_i] = \phi_{MLP}\big([\boldsymbol{\theta}_i \parallel \phi_\tau(\hat{\tau}_i^*)]\big), \quad (44)$$

where A_i controls the interaction intensity and σ_i determines its spatial extent.

Definition (Social Utility)**.** *For a candidate ego trajectory τ_e, the social utility evaluates its compatibility with neighboring agents under the agent-conditioned interaction fields:*

$$\mathcal{U}_s(\tau_e) = \exp\left(-\max_{t\in[0,T]} \sum_{i=1}^N A_i \exp\left(-\frac{\|p_e(t) - \hat{p}_i^*(t)\|^2}{2\sigma_i^2}\right)\right). \quad (45)$$

Here, $\hat{p}_i^*(t)$ denotes the predicted position of neighboring agent i at time t. The resulting utility jointly depends on geometric proximity, behavioral objective coordinates, and predicted future motions.

To balance ego-driving quality and social accommodation, we introduce a continuous SVO angle $\varphi \in [0, \pi/2]$. Smaller φ emphasizes ego-driving quality, while larger φ places greater weight on social accommodation. After monotone normalization of $\mathcal{U}_e$ and $\mathcal{U}_s$, the interaction utility is defined as

$$\mathcal{U}_b(\tau_e) = \cos(\varphi)\tilde{\mathcal{U}}_e(\tau_e) + \sin(\varphi)\tilde{\mathcal{U}}_s(\tau_e). \quad (46)$$

A.6 Monotonicity and Curvature Characterization

Theorem 2 (Monotonicity and Curvature Characterization)**.** *Let $\mathcal{V}(u;\mathbf{w}) = \sum_{q=0}^d w_q b_{q,d}(u)$ be defined on $u \in [0,1]$, with $w_0 = 0$ and*

$$w_q = \sum_{j=1}^q \delta_j, \qquad q = 1, \dots, d. \quad (47)$$

If $\delta_j > 0$ for all $j = 1, \dots, d$, then $\mathcal{V}$ is strictly increasing on $[0,1]$, i.e., $\mathcal{V}'(u) > 0$ for all $u \in [0,1]$. Moreover, if $\delta_{q+2} - \delta_{q+1} \le 0$ for all $q = 0, \dots, d-2$, then $\mathcal{V}$ is concave on $[0,1]$; if $\delta_{q+2} - \delta_{q+1} \ge 0$ for all $q = 0, \dots, d-2$, then $\mathcal{V}$ is convex on $[0,1]$.

Proof. For a Bernstein polynomial of degree d, the first and second derivatives satisfy

$$\mathcal{V}'(u) = d\sum_{q=0}^{d-1}(w_{q+1} - w_q) b_{q,d-1}(u) \quad (48)$$

and

$$\mathcal{V}''(u) = d(d-1)\sum_{q=0}^{d-2}(w_{q+2} - 2w_{q+1} + w_q) b_{q,d-2}(u). \quad (49)$$

Under Eq. 16,

$$w_{q+1} - w_q = \delta_{q+1}, \tag{50}$$

and

$$w_{q+2} - 2w_{q+1} + w_q = \delta_{q+2} - \delta_{q+1}. \tag{51}$$

Substituting into Eq. 48 gives

$$\mathcal{V}'(u) = d \sum_{q=0}^{d-1} \delta_{q+1} b_{q,d-1}(u). \tag{52}$$

Since $\delta_{q+1} > 0$ and the Bernstein basis functions satisfy $b_{q,d-1}(u) \geq 0$ and $\sum_{q=0}^{d-1} b_{q,d-1}(u) = 1$ on $[0,1]$, it follows that $\mathcal{V}'(u) > 0$ for all $u \in [0,1]$. Hence, $\mathcal{V}$ is strictly increasing.

Similarly,

$$\mathcal{V}''(u) = d(d-1) \sum_{q=0}^{d-2} (\delta_{q+2} - \delta_{q+1}) b_{q,d-2}(u). \tag{53}$$

Since $b_{q,d-2}(u) \geq 0$ on $[0,1]$, we obtain

$$\delta_{q+2} - \delta_{q+1} \leq 0, \ \forall q \quad \Longrightarrow \quad \mathcal{V}''(u) \leq 0, \ \forall u \in [0,1], \tag{54}$$

and

$$\delta_{q+2} - \delta_{q+1} \geq 0, \ \forall q \quad \Longrightarrow \quad \mathcal{V}''(u) \geq 0, \ \forall u \in [0,1]. \tag{55}$$

Therefore, $\mathcal{V}$ is concave in the former case and convex in the latter case, while more general increment sequences allow context-dependent changes in curvature. □

A.7 Context-Adaptive Value Planning

Both the SVO angle φ and the value-function parameters $\mathbf{w}$ should adapt to the ongoing interaction rather than remain fixed across scenarios. We therefore infer them from surrounding-agent representations and predicted futures, and use them to construct context-dependent decision representations for ego planning.

For each neighboring agent $i \in \{1, \ldots, N\}$, we encode its interaction state as

$$h_i^{int} = \phi_{MLP}\big([F_i \parallel \boldsymbol{\theta}_i \parallel \phi_\tau(\hat{\tau}_i^*)]\big), \tag{56}$$

where F_i denotes the contextual feature of agent i, $\boldsymbol{\theta}_i$ denotes its behavioral objective coordinates, $\hat{\tau}_i^*$ is the selected predicted trajectory, and ϕ_τ is a trajectory encoder. We aggregate these interaction tokens through ego-centered attention:

$$F_{int} = \phi_{Attn}\big(Q = F_e, \ K = \{h_i^{int}\}_{i=1}^N, \ V = \{h_i^{int}\}_{i=1}^N\big). \tag{57}$$

The resulting interaction context summarizes the predicted motions of surrounding agents together with the structured behavioral information relevant to ego interaction.

Based on the ego representation F_e and interaction context F_{int}, a lightweight hyper-network ϕ_{HN} generates the context-dependent SVO and value-function parameters:

$$\begin{aligned} [\tilde{\varphi} \parallel \mathbf{z}_\delta] &= \phi_{HN}([F_e \parallel F_{int}]) \\ &= \phi_{MLP}^{(2)}\Big(\phi_{SiLU}\Big(\phi_{LN}\Big(\phi_{MLP}^{(1)}([F_e \parallel F_{int}])\Big)\Big)\Big). \end{aligned} \tag{58}$$

Here, $\tilde{\varphi} \in \mathbb{R}$ is the SVO logit and $\mathbf{z}_\delta \in \mathbb{R}^d$ contains the logits of the d positive increments.

The SVO angle is obtained through

$$\varphi = \frac{\pi}{2} \sigma(\tilde{\varphi}), \tag{59}$$

while the increment sequence is parameterized as

$$\delta_q = \ln(1 + e^{(\mathbf{z}_\delta)_q}), \qquad q = 1, \ldots, d. \tag{60}$$

The value-function control points are then reconstructed by

$$w_0 = 0, \qquad w_q = \sum_{j=1}^{q} \delta_j, \qquad q = 1, \dots, d. \tag{61}$$

By construction, the positive increments guarantee that the resulting value function is strictly monotone, while their relative variation determines its curvature according to Theorem 2. Thus, φ controls the balance between ego and social utility, whereas $\mathbf{w}$ controls the context-dependent transformation from interaction utility to decision value.

We next project these parameters into a unified decision embedding. Let ϕ_φ and ϕ_w denote projection networks for the SVO angle and value-function control points, respectively, and let ϕ_{Gate} fuse the two representations:

$$\mathcal{E}_\psi = \phi_{Gate}\big([\phi_\varphi(\varphi) \parallel \phi_w(\mathbf{w})]\big). \tag{62}$$

Given the interaction context F_{int} and decision embedding $\mathcal{E}_\psi$, we construct the ego planning query. The ego representation is first projected as

$$Q_e = \phi_L(F_e), \tag{63}$$

and then fused with the decision embedding and interaction context:

$$Q_{plan} = \phi_{MLP}\big([Q_e \parallel \mathcal{E}_\psi \parallel F_{int}]\big). \tag{64}$$

Global scene information is further incorporated through cross-attention:

$$Q_{attn} = \phi_{Attn}(Q = Q_{plan},\ K = F_c,\ V = F_c). \tag{65}$$

Conditioned on Q_{attn}, a GRU-based decoder generates K candidate ego trajectories and their mode probabilities:

$$\{(\hat{\tau}_e^k, \hat{\pi}_e^k)\}_{k=1}^K = \phi_{GRU}(Q_{attn}). \tag{66}$$

The resulting multimodal ego trajectory set is $\hat{\mathcal{T}}_e = \{\hat{\tau}_e^k\}_{k=1}^K$. Ego planning is therefore conditioned jointly on scene context, neighboring predicted futures, behavioral objective coordinates, and context-adaptive valuation parameters.

B Training Objective

B.1 Trajectory Prediction Loss

The prediction objective $\mathcal{L}_{traj}$ accounts for the multimodal nature of surrounding-agent forecasting. It consists of a regression loss $\mathcal{L}_{reg}^{pred}$ for trajectory prediction, a classification loss $\mathcal{L}_{cls}^{pred}$ for mode assignment, and a minimum Average Displacement Error term $\mathcal{L}_{ade}$:

$$\mathcal{L}_{traj} = \mathcal{L}_{ade} + \mathcal{L}_{reg}^{pred} + \mathcal{L}_{cls}^{pred}. \tag{67}$$

Here, $\mathcal{L}_{reg}^{pred}$ and $\mathcal{L}_{cls}^{pred}$ follow the standard multimodal supervision protocol.

To learn the behavioral objective coordinates, we further introduce a MaxEnt-inspired contrastive ranking objective. For the demonstrated future trajectory τ^* of agent i and its optimized multimodal prediction set $\hat{\mathcal{T}}^i = \{\hat{\tau}_i^k\}_{k=1}^K$, we define

$$\mathcal{L}_{coord} = -\log \frac{\exp\left(R(\tau^*; \boldsymbol{\theta}_i)\right)}{\exp\left(R(\tau^*; \boldsymbol{\theta}_i)\right) + \sum_{k=1}^K \exp\left(R(\hat{\tau}_i^k; \boldsymbol{\theta}_i)\right)}. \tag{68}$$

This objective encourages the demonstrated future to receive a higher behavioral trajectory score than the generated alternatives, thereby learning agent-conditioned coordinates that are consistent with the observed interaction behavior.

The total prediction loss is

$$\mathcal{L}_{pred} = \mathcal{L}_{traj} + \lambda_c \mathcal{L}_{coord}, \tag{69}$$

where λ_c balances geometric forecasting and behavioral coordinate learning.

B.2 Trajectory Planning Loss

The planning stage is trained through geometric imitation and contextual value alignment. Similar to the prediction stage, the planner produces multimodal ego-trajectory hypotheses together with their mode probabilities. The imitation loss is defined as

$$\mathcal{L}_{imit} = \mathcal{L}_{reg}^{plan} + \mathcal{L}_{cls}^{plan}. \tag{70}$$

To jointly learn context-dependent valuation and ego planning, we introduce a contrastive value-alignment objective. Let

$$u(\tau_e) = \mathrm{Norm}\left(\mathcal{U}_b(\tau_e)\right) \in [0, 1]. \tag{71}$$

For the demonstrated ego trajectory τ_e^* and generated alternatives $\{\hat{\tau}_e^k\}_{k=1}^K$, the value-alignment loss is defined as

$$\mathcal{L}_{value} = -\log \frac{\exp\left(\mathcal{V}(u(\tau_e^*); \mathbf{w})\right)}{\exp\left(\mathcal{V}(u(\tau_e^*); \mathbf{w})\right) + \sum_{k=1}^K \exp\left(\mathcal{V}(u(\hat{\tau}_e^k); \mathbf{w})\right)}. \tag{72}$$

This objective encourages the demonstrated trajectory to attain a higher context-dependent decision value than the generated alternatives under the learned Prospect Value Model.

The total planning loss is

$$\mathcal{L}_{plan} = \mathcal{L}_{imit} + \lambda_v \mathcal{L}_{value}, \tag{73}$$

where λ_v controls the contribution of contextual value alignment.

The overall training objective is

$$\mathcal{L} = \mathcal{L}_{pred} + \mathcal{L}_{plan}. \tag{74}$$

Both objectives follow the same demonstration-ranking principle: surrounding-agent coordinates are learned by ranking observed futures in the behavioral objective space, while ego valuation is learned by ranking expert plans in the decision-value space.

C Experimental Details

C.1 Experimental Setup

C.1.1 Benchmarks

We evaluate the proposed model on five benchmarks. Specifically, nuScenes, NGSIM, and HighD are used as trajectory-prediction datasets, while nuPlan and CommonRoad are used as planning benchmarks.

nuScenes. nuScenes (Caesar et al., 2020) is a large-scale urban benchmark collected in Boston and Singapore, featuring complex intersections and dense traffic. Following standard protocols, we use a 2 s history to predict a 6 s future horizon. The dataset is split into 700, 150, and 150 scenes for training, validation, and testing, respectively.

NGSIM. NGSIM (Deo & Trivedi, 2018) was captured from US highways and focuses on dense, naturalistic traffic flows with frequent complex maneuvers. Raw data is downsampled to 5 Hz. We use a 3 s history to forecast the subsequent 5 s, with a chronological split of 70%/10%/20%.

HighD. HighD (Krajewski et al., 2018) contains over 110,000 vehicle motions recorded by drones on German highways, providing high-precision trajectories with minimal occlusions. Similar to NGSIM, we process the data at 5 Hz and predict a 5 s future based on a 3 s history.

nuPlan. The nuPlan dataset (Caesar et al., 2021) is a large-scale benchmark from which we select around 1 million scenarios for training. For closed-loop evaluation, we employ the Val14, Test14, and Test14-hard subsets, with each scenario simulated at 10 Hz for 15 s, featuring high-definition maps and expert-annotated ego trajectories.

CommonRoad. CommonRoad is a motion-planning benchmark containing partly real-world recorded and partly handcrafted scenarios, including safety-critical situations across urban, rural, and highway environments.

C.1.2 Evaluation Metrics

For trajectory prediction on nuScenes, we report standard multimodal forecasting metrics, including Minimum Average Displacement Error (minADE_K), Minimum Final Displacement Error (minFDE_K), and Miss Rate (MR_K). For NGSIM and HighD, we report Root Mean Square Error (RMSE). For trajectory planning on nuPlan, we follow the official closed-loop evaluation protocol and report the overall closed-loop score together with sub-metrics including no at-fault collisions, time-to-collision within bound, drivable area compliance, comfort, progress, and speed-limit compliance. For CommonRoad, DIST denotes the average route-completion distance, JERK denotes the average jerk cost, and RC denotes the reaction cost induced by nearby agents within a 40-m range. CT counts valid ego collisions excluding rear-end collisions and collisions with stationary agents, while RCT counts rear-end collisions with following agents.

C.1.3 Implementation details

HSP is implemented in PyTorch and trained on eight NVIDIA A100 GPUs. We use the Adam optimizer with an initial learning rate of 10^{-3} and a cosine annealing warm-restart scheduler. The model is trained for 25 epochs with a batch size of 128, and the learning rate is decayed by a factor of 2 at each restart stage. The number of multimodal trajectory hypotheses is set to $K = 12$, and the degree of the Bernstein polynomial in the Prospect Value Model is set to $d = 4$. The weighting coefficients of the coordinate-learning loss and value-alignment loss are set to $\lambda_c = 0.5$ and $\lambda_v = 0.2$, respectively.

C.2 Additional Prediction and Planning Results

C.2.1 Additional Prediction Results

The results on NGSIM and HighD in Table A.1 further demonstrate the effectiveness of HSP in highway traffic. On NGSIM, HSP achieves the best performance across all prediction horizons and improves the average RMSE by 12.4% compared with the strongest competing method. The advantage becomes more pronounced at longer horizons, where the 5 s RMSE drops from 2.87 to 2.36, indicating that the proposed model is better able to capture the cumulative effect of interaction behaviors over time. A similar trend is observed on HighD. HSP improves the average RMSE by 41.4%, and at the 5 s horizon, the RMSE is reduced from 0.61 to 0.31. The margin generally increases toward longer horizons, indicating stronger long-horizon forecasting performance across highway scenarios.

Table A.1: Comparison of model performance on the NGSIM and HighD datasets. Metric: RMSE.

Dataset	Model	Prediction Horizon (s)					
		1	2	3	4	5	AVG
NGSIM	CF-LSTM (Xie et al., 2021)	0.55	1.10	1.78	2.73	3.82	2.00
	iNATran (Chen et al., 2022a)	0.39	0.96	1.61	2.42	3.43	1.76
	STDAN (Chen et al., 2022b)	0.39	0.96	1.62	2.51	3.65	1.83
	WSiP (Wang et al., 2023b)	0.56	1.23	2.05	3.08	4.34	2.25
	HTPF (Wang et al., 2023a)	0.49	1.09	1.78	2.62	3.65	1.92
	HiT (Liao et al., 2025a)	0.38	0.90	<u>1.42</u>	<u>2.08</u>	<u>2.87</u>	<u>1.53</u>
	DEMO (Wang et al., 2025c)	<u>0.36</u>	<u>0.86</u>	1.48	2.10	2.88	1.54
	HSP (Ours)	**0.36**	**0.84**	**1.35**	**1.78**	**2.36**	**1.34**
HighD	CF-LSTM (Xie et al., 2021)	0.18	0.42	1.07	1.72	2.44	1.17
	STDAN (Chen et al., 2022b)	0.19	0.27	0.48	0.91	1.66	0.70
	DRBP (Gao et al., 2023)	0.41	0.79	1.11	1.40	2.58	1.26
	WSiP (Wang et al., 2023b)	0.20	0.60	1.21	2.07	3.14	1.44
	HTPF (Wang et al., 2023a)	0.16	0.47	0.94	1.58	2.36	1.10
	GaVa (Liao et al., 2024b)	0.17	0.24	0.42	0.86	1.31	0.60
	HiT (Liao et al., 2025a)	0.08	<u>0.13</u>	<u>0.22</u>	<u>0.39</u>	<u>0.61</u>	<u>0.29</u>
	DEMO (Wang et al., 2025c)	**0.06**	0.14	0.25	0.44	0.70	0.32
	HSP (Ours)	<u>0.07</u>	**0.12**	**0.16**	**0.20**	**0.31**	**0.17**

C.2.2 Additional Planning Results

Table A.2 reports additional closed-loop planning results on the *Test14-hard* and *Test14* splits of nuPlan. HSP ranks first under both reactive and non-reactive settings on the harder *Test14-hard* split, and achieves the best reactive score while remaining highly competitive under the non-reactive setting on the full *Test14* split. The stronger advantage on *Test14-hard* suggests that HSP is particularly effective when interaction uncertainty and decision difficulty increase.

Table A.2: Closed-loop planning results on the *Test14-hard* and *Test14* splits of the nuPlan benchmark. We report the overall closed-loop score under both CLS-R and CLS-NR settings. Cases marked with ('-') indicate missing values.

Planner	Test14-hard		Test14	
	CLS-R↑	CLS-NR↑	CLS-R↑	CLS-NR↑
Log-replay	68.80	85.96	75.86	94.03
GC-PGP (Hallgarten et al., 2023)	39.63	43.22	51.39	55.99
RasterModel (Caesar et al., 2021)	52.16	49.47	67.54	69.66
PDM-Open (Dauner et al., 2023)	35.83	33.51	57.23	52.80
UrbanDriver (Scheel et al., 2022)	49.95	50.40	61.02	63.27
PDM-Closed (Dauner et al., 2023)	75.19	65.08	<u>91.64</u>	90.05
PDM-Hybrid (Dauner et al., 2023)	<u>76.07</u>	65.99	91.28	90.10
GameFormer (Huang et al., 2023a)	67.05	68.70	79.31	80.80
PEP (Zhang et al., 2024a)	-	-	89.74	91.45
IDM (Wong et al., 2024)	62.26	56.15	72.42	70.39
PlanTF (Cheng et al., 2024b)	61.61	<u>69.70</u>	80.59	86.48
CarPlanner (Zhang et al., 2025)	-	-	91.10	**94.07**
HSP (Ours)	**77.03**	**82.18**	**91.86**	<u>93.63</u>

Table A.3 reports closed-loop planning results on CommonRoad, which contains interactive scenarios across different road geometries and traffic configurations. HSP achieves the strongest performance across all reported metrics, obtaining the largest route-completion distance while reducing jerk, reaction cost, and both collision measures. These results indicate that the planning gains observed on nuPlan extend to a different benchmark and evaluation protocol, supporting the cross-benchmark robustness of the proposed prediction–planning framework.

Table A.3: Closed-loop simulation results on CommonRoad. The bold values indicate the best performance, and underlined values indicate the second-best performance.

Model	DIST↑	JERK↓	RC↓	CT↓	RCT↓
DIPP (Huang et al., 2023b)	14.71	6.78	0.69	19	15
GameFormer (Huang et al., 2023a)	16.77	11.29	<u>0.65</u>	20	12
HiVT (+Planner) (Zhou et al., 2022)	<u>19.54</u>	<u>4.44</u>	0.66	16	<u>8</u>
PGP (+Planner) (Deo et al., 2022)	19.43	7.71	0.65	<u>13</u>	9
HSP (Ours)	**20.85**	**3.12**	**0.62**	**9**	**6**

C.3 Computational Efficiency

We provide additional analysis of computational efficiency for both trajectory prediction and ego planning. On nuScenes, HSP requires 20 ms to predict 12 surrounding agents and achieves the strongest forecasting performance among the compared methods in Table A.4. For planning, HSP uses 12.9 G FLOPs and 19.0 M parameters while achieving a CLS-NR score of 93.33 on nuPlan, yielding a favorable performance–complexity trade-off compared with the evaluated baselines (Table A.5). These results indicate that the proposed structured behavioral modeling and contextual valuation do not rely on increased model scale to obtain their prediction and planning gains.

Table A.4: Comparative analysis of prediction inference efficiency on nuScenes.

Model	Inference Time (ms)	minADE_5	minFDE_1	MR_5
MultiPath (Chai et al., 2019)	87	1.44	7.69	0.74
Trajectron++ (Salzmann et al., 2020)	38	1.88	9.52	0.70
AgentFormer (Yuan et al., 2021)	107	1.97	9.12	0.69
LAformer (Liu et al., 2024)	115	1.19	6.89	0.54
VisionTrap (Moon et al., 2024)	53	1.36	8.72	0.61
HSP (Ours)	**20**	**1.17**	**6.54**	**0.49**

Table A.5: Comparison of model complexity and planning performance on nuPlan.

Model	FLOPs (G)	Params (M)	CLS-NR↑
UrbanDriver (Scheel et al., 2022)	28.5	35.2	68.57
PDM-Open (Dauner et al., 2023)	32.9	41.4	53.53
GC-PGP (Hallgarten et al., 2023)	65.2	79.3	61.09
GameFormer (Huang et al., 2023a)	85.6	95.3	82.95
PlanTF (Cheng et al., 2024b)	90.2	149.6	85.30
HSP (Ours)	**12.9**	**19.0**	**93.33**

C.4 Additional Ablation and Representation Analysis

C.4.1 Prediction Components.

Table A.6 evaluates the three components that operationalize the behavioral objective coordinates in trajectory prediction: objective-structured decoding (OSD), coordinate learning (CL), and energy-driven trajectory optimization (EDTO). Removing any component consistently degrades forecasting performance. OSD has the largest individual effect, showing that explicitly conditioning multimodal generation on the learned coordinates is important for prediction. CL further improves the quality of the inferred coordinates, while EDTO provides an additional gain by directly refining candidate futures under the coordinate-conditioned trajectory energy.

Table A.6: Ablation of prediction-stage components on nuScenes. OSD: Objective-Structured Decoding; CL: Coordinate Learning; EDTO: Energy-Driven Trajectory Optimization.

Model	Components			Performance		
	OSD	CL	EDTO	minADE_5	minFDE_5	MR_5
A	×	✓	✓	1.28	2.41	0.57
B	✓	×	✓	1.25	2.39	0.55
C	✓	✓	×	1.23	2.34	0.52
D	✓	✓	✓	**1.17**	**2.26**	**0.49**

C.4.2 Dimension-Wise Objective Ablations

To examine the distinct roles of the four behavioral objectives, we remove one objective axis at a time while keeping the remaining architecture and training settings unchanged. As shown in Table A.7, removing any objective degrades both prediction and planning performance. More importantly, the downstream effects are semantically aligned with the removed axis: safety primarily affects collision and TTC, comfort most strongly affects the comfort score, efficiency has the largest impact on progress and speed, and social accommodation mainly affects drivable-area compliance. These results indicate that the four coordinates provide complementary and behaviorally structured information rather than acting as interchangeable latent features.

Table A.7: Dimension-wise ablation of behavioral objective coordinates on nuScenes and nuPlan *Test14-hard* split.

	nuScenes			nuPlan *Test14-hard*						
Variant	minADE_5 ↓	minFDE_5 ↓	MR_5 ↓	CLS-R↑	Collision↑	TTC↑	Drivable↑	Comfort↑	Progress↑	Speed↑
w/o Safety	1.19	2.33	0.50	74.58	90.92	83.27	97.94	86.73	71.84	98.18
w/o Comfort	1.18	2.28	0.49	76.51	95.08	87.91	98.39	82.46	71.81	98.51
w/o Efficiency	1.20	2.32	0.51	74.95	92.37	85.14	97.98	85.91	69.06	96.81
w/o Social Accommodation	1.19	2.29	0.50	75.64	93.18	85.96	97.61	86.68	71.42	98.13
Full HSP	**1.17**	**2.26**	**0.49**	**77.03**	**95.34**	**88.43**	**98.51**	**87.31**	**72.57**	**98.56**

C.4.3 Coordinate-Space Structure

To examine the internal structure of the learned behavioral objective coordinates, we project them into a two-dimensional space using UMAP, as shown in Fig. A.1. The resulting embedding exhibits several coherent regions rather than an unstructured scatter, indicating that the learned coordinates preserve meaningful organization across heterogeneous interaction patterns. We further visualize

the normalized intensity of each coordinate on the same embedding. The four axes—safety, comfort, efficiency, and social accommodation—each form distinct yet spatially smooth gradients. This suggests that the learned representation does not merely encode arbitrary hidden factors, but organizes agents according to structured behavioral variation along interpretable objective axes. At the same time, partial overlap among the four gradients indicates that these objectives are not mutually exclusive, which is consistent with the coupled nature of real driving behavior. Overall, this visualization provides qualitative evidence that the learned coordinates form a structured and behaviorally meaningful representation space beyond an unconstrained latent embedding.

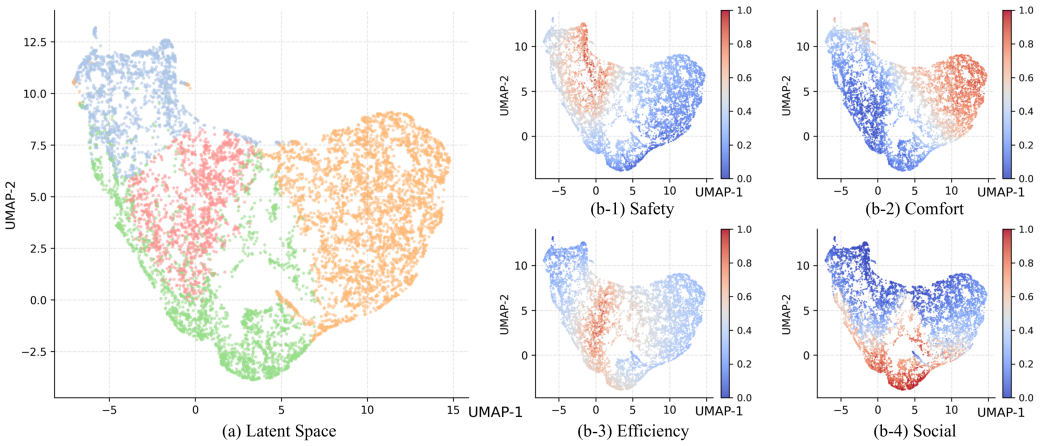

Figure A.1: Structure of the learned behavioral objective coordinates on nuScenes. Panel (a) shows a UMAP projection of the learned coordinates, with colors indicating K-Means clusters. Panels (b-1)–(b-4) show the same embedding colored by the normalized safety, comfort, efficiency, and social-accommodation coordinates.

C.4.4 EDTO Refinement and Feasibility

We further analyze the role of Energy-Driven Trajectory Optimization (EDTO) beyond its contribution to forecasting accuracy. EDTO improves endpoint prediction accuracy while preserving trajectory feasibility. As shown in Table A.8, refinement reduces minFDE_5 from 2.34 to 2.26, lowers the collision rate from 2.63% to 2.51%, and slightly reduces RMS jerk from 2.58 to 2.54. The off-road rate remains nearly unchanged (1.21% vs. 1.23%), indicating that gradient-based refinement improves prediction quality without materially degrading map compliance or motion smoothness.

Table A.8: Effect of EDTO refinement on prediction accuracy and trajectory feasibility.

Variant	$\text{minFDE}_5 \downarrow$	Off-road↓	Collision↓	RMS Jerk↓
Initial Proposal	2.34	**1.21%**	2.63%	2.58
+ EDTO	**2.26**	1.23%	**2.51%**	**2.54**

C.5 Additional Prediction–Utility Analysis

We further examine the relationship between geometric prediction accuracy and downstream planning performance by grouping scenarios by prediction ADE and visualizing planning-score distributions within each bin. As shown in Fig. A.2, planning performance does not vary monotonically with prediction error. In particular, scenarios within similar low- and medium-ADE ranges exhibit substantial variation in planning outcomes, and low geometric error does not necessarily correspond to high planning scores. This indicates that displacement-based prediction metrics capture only part of the information relevant to downstream decision making: predictions with similar geometric accuracy can still differ in how they characterize interaction-relevant future behavior.

To control for geometric forecasting quality more directly, the main paper further compares HSP with a geometry-driven predictor–planner baseline under the matched-accuracy criterion $|\Delta\mathrm{ADE}| \leq 0.1$ m. Under this constraint, the two pipelines have comparable displacement accuracy, while HSP retains positive mean planning-score gains of 0.20 on *Val14* and 0.39 on *Test14-hard*. The stronger gain on the harder split indicates that the distinction between geometric accuracy and decision relevance becomes more pronounced in challenging interactive scenarios. Together, the two analyses provide complementary evidence for the prediction–utility gap: geometric forecasting quality alone does not fully determine the downstream value of a predicted future.

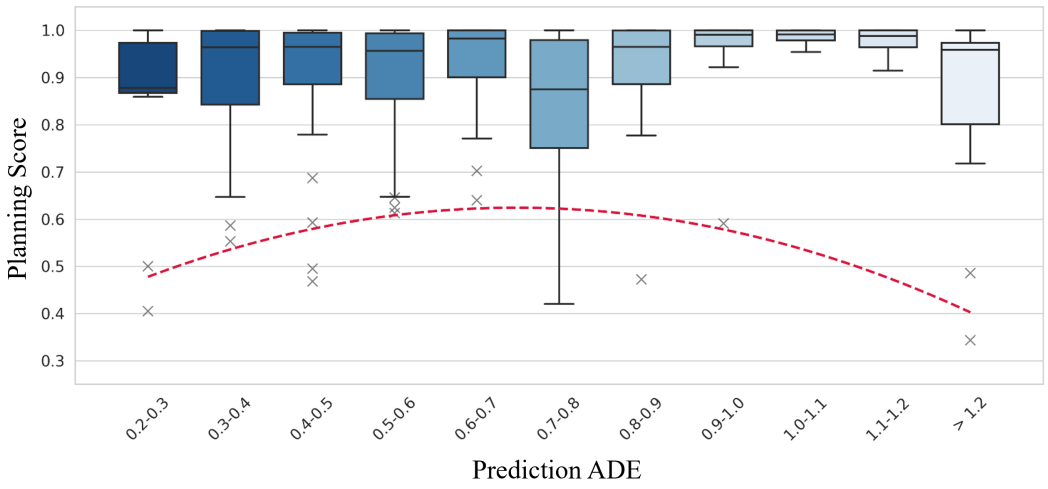

Figure A.2: Relationship between geometric prediction accuracy and downstream planning performance. Scenarios are grouped by prediction ADE, with the distribution of planning scores shown for each bin. Substantial planning-score variation remains within similar ADE ranges, indicating that geometric forecasting accuracy alone does not fully characterize decision relevance.

C.6 Additional Planning Visualizations

Fig. A.3 visualizes the closed-loop behavior of HSP in four representative nuPlan scenarios, including lane changing, turning, intersection negotiation, and roundabout driving. Across these cases, the ego vehicle evaluates multiple candidate trajectories and adapts its final plan to the evolving motions of surrounding vehicles and pedestrians. The examples illustrate how objective-structured predictions and context-adaptive interaction valuation jointly support planning responses that balance progress with interaction safety, without systematically resorting to conservative stopping.

Figs. A.4 and A.5 provide additional qualitative results on CommonRoad. Fig. A.4 covers intersection crossing, lane merging, and dense highway navigation, while Fig. A.5 focuses on three intersection interaction scenarios. Across these cases, HSP adapts the ego trajectory to predicted surrounding-agent motions and changing interaction geometry. The examples show that the planner can negotiate conflict regions while maintaining forward progress, illustrating consistent interaction-aware planning behavior across different driving environments.

D Limitations and Discussion

Limitations. HSP provides a structured framework for connecting behavioral representation, trajectory prediction, and interaction-aware planning in mixed traffic. The four behavioral objectives—safety, comfort, efficiency, and social accommodation—form a task-grounded inductive basis rather than an exhaustive characterization of driving behavior; additional objectives or alternative parameterizations may be needed for different traffic environments and road-user populations. Driving behavior also varies across countries, regions, and traffic cultures, potentially introducing distribution shifts that require domain-specific calibration and validation. In addition, the Prospect Value Model

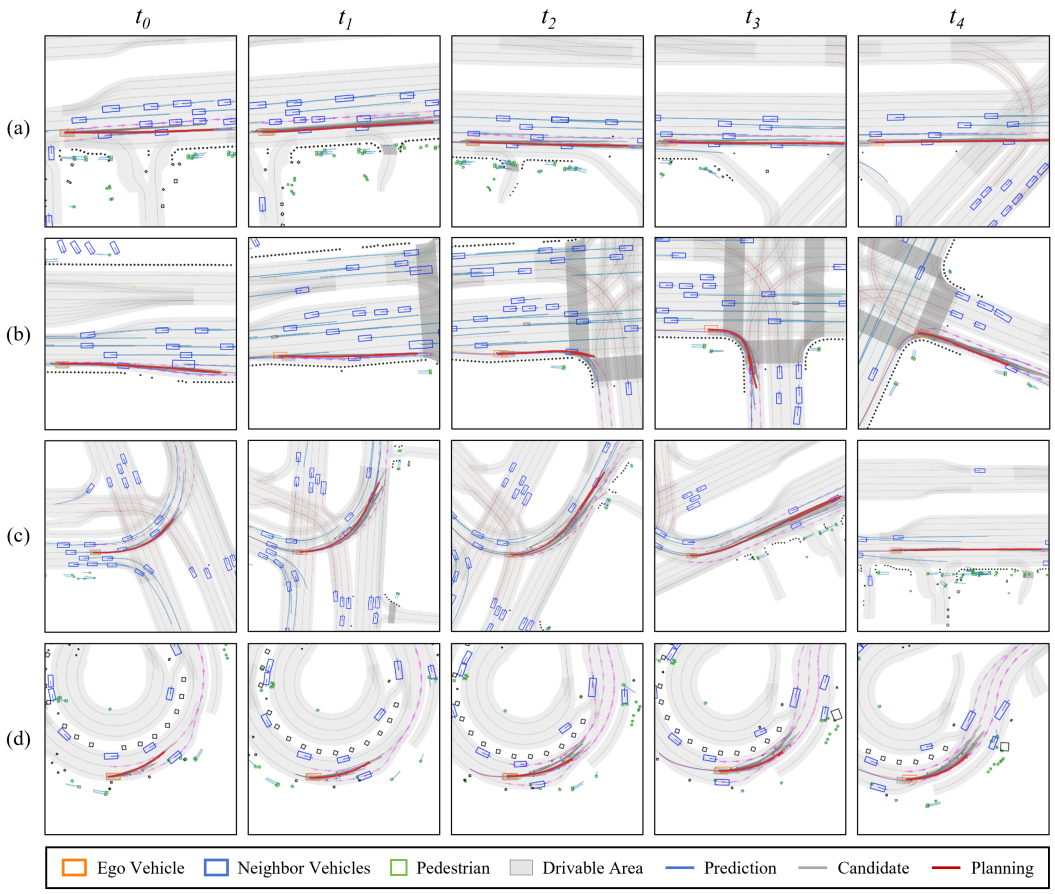

Figure A.3: Closed-loop planning examples on nuPlan: (a) straight-road lane changing, (b) right-turn with lane change, (c) intersection left-turn, and (d) roundabout navigation. From t_0 to t_4, the ego vehicle (orange) evaluates candidate trajectories (gray) and generates a final plan (red) in response to predicted motions (blue lines) of surrounding vehicles (blue) and pedestrians (green). Planning is conditioned on objective-structured predictions and context-adaptive interaction valuation.

is motivated by nonlinear decision valuation rather than intended to identify canonical prospect-theory parameters. Finally, benchmark improvements do not constitute real-world safety validation, and learned interaction valuation should operate together with independent hard-safety constraints, out-of-distribution detection, and conservative fallback mechanisms.

Broader Impacts. HSP may contribute to more transparent and interaction-aware autonomous-driving systems by exposing structured behavioral objective coordinates rather than relying solely on implicit prediction features. Because the same representation is used across forecasting, trajectory refinement, and downstream planning, it provides an explicit interface for analyzing how interaction-relevant behavioral information affects autonomous decisions. The context-adaptive value model further offers a way to study how nonlinear valuation influences planning under different interaction conditions. More broadly, such structured prediction–planning interfaces may support future research on interpretable, socially aware, and safety-conscious autonomous decision making.

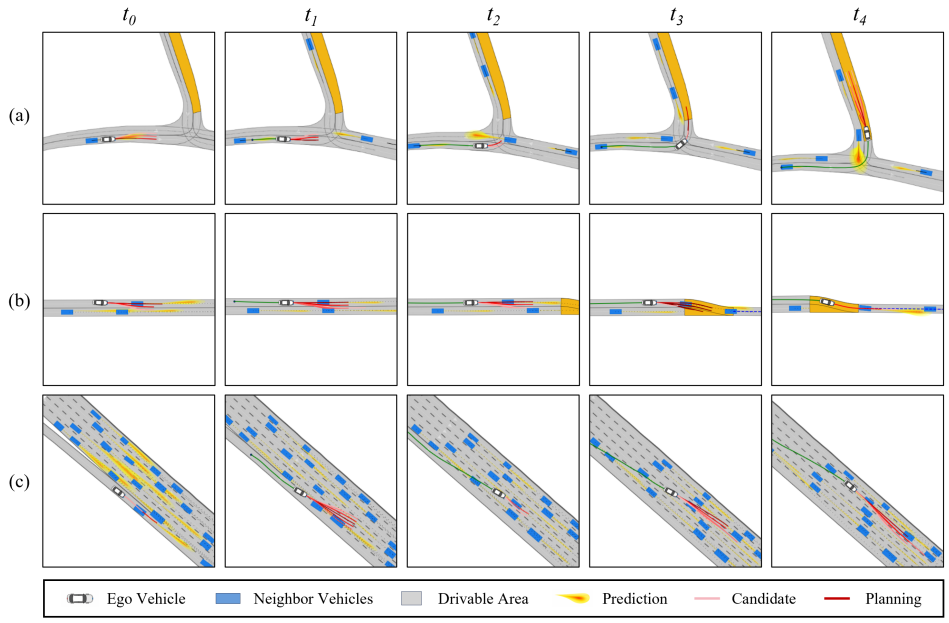

Figure A.4: Qualitative planning results on CommonRoad across three interactive scenarios: (a) intersection crossing, (b) lane merging, and (c) dense highway navigation.

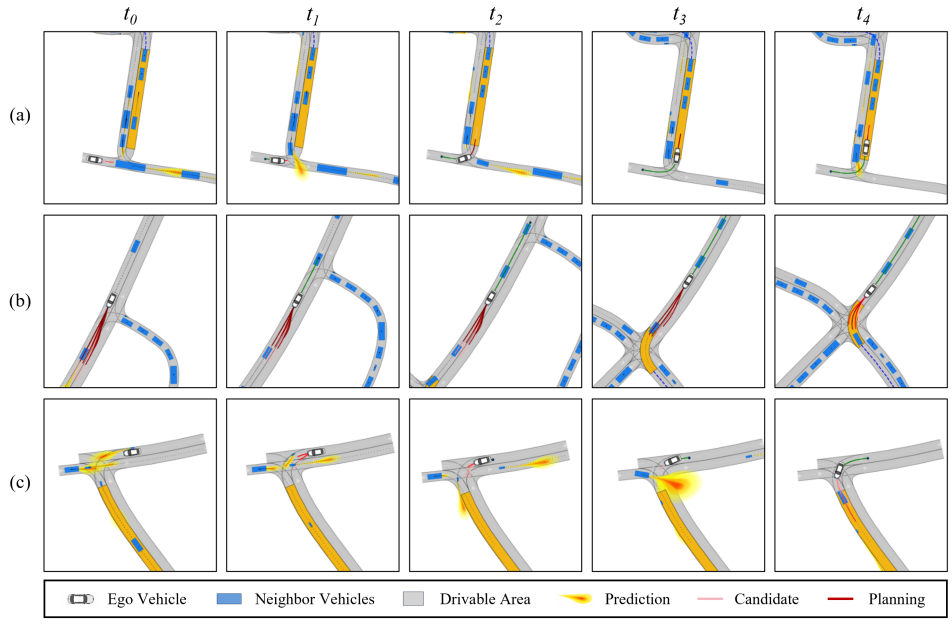

Figure A.5: Qualitative planning results on CommonRoad across three intersection interaction scenarios. Panels (a)–(c) show different conflict configurations requiring adaptive ego responses to surrounding traffic.