Social behavior for autonomous vehicles

Schwarting, Wilko; Pierson, Alyssa; Alonso-Mora, Javier; Karaman, Sertac; Rus, Daniela

DOI
10.1073/pnas.1820676116

Publication date
2019

Document Version
Final published version

Published in
Proceedings of the National Academy of Sciences of the United States of America

Social behavior for autonomous vehicles

Wilko Schwarting[a,1], Alyssa Pierson[a], Javier Alonso-Mora[b], Sertac Karaman[c], and Daniela Rus[a]

[a]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139; [b]Cognitive Robotics, Delft University of Technology, 2628 CD, Delft, Netherlands; and [c]Laboratory of Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by Paul A. M. Van Lange, VU Amsterdam, Amsterdam, Netherlands, and accepted by Editorial Board Member Susan T. Fiske October 24, 2019 (received for review December 6, 2018)

Deployment of autonomous vehicles on public roads promises increased efficiency and safety. It requires understanding the intent of human drivers and adapting to their driving styles. Autonomous vehicles must also behave in safe and predictable ways without requiring explicit communication. We integrate tools from social psychology into autonomous-vehicle decision making to quantify and predict the social behavior of other drivers and to behave in a socially compliant way. A key component is Social Value Orientation (SVO), which quantifies the degree of an agent's selfishness or altruism, allowing us to better predict how the agent will interact and cooperate with others. We model interactions between agents as a best-response game wherein each agent negotiates to maximize their own utility. We solve the dynamic game by finding the Nash equilibrium, yielding an online method of predicting multiagent interactions given their SVOs. This approach allows autonomous vehicles to observe human drivers, estimate their SVOs, and generate an autonomous control policy in real time. We demonstrate the capabilities and performance of our algorithm in challenging traffic scenarios: merging lanes and unprotected left turns. We validate our results in simulation and on human driving data from the NGSIM dataset. Our results illustrate how the algorithm's behavior adapts to social preferences of other drivers. By incorporating SVO, we improve autonomous performance and reduce errors in human trajectory predictions by 25%.

autonomous driving | Social Value Orientation | social compliance | game theory | inverse reinforcement learning

Interacting with human drivers is one of the great challenges of autonomous driving. To operate in the real world, autonomous vehicles (AVs) need to cope with situations requiring complex observations and interactions, such as highway merging and unprotected left-hand turns, which are challenging even for human drivers. For example, over 450,000 lane-change/merging accidents and 1.4 million right-/left-turn accidents occurred in the United States in 2015 alone (1). Currently, AVs lack an understanding of human behavior, thus requiring conservative behavior for safe operation. Conservative driving creates bottlenecks in traffic flow, especially in intersections. For example, Waymo, considered a leader in autonomous driving, still struggles with left turns and acting in predictable manners (2). This conservative behavior not only leaves AVs vulnerable to aggressive human drivers and inhibits the interpretability of intentions, but also can result in unexpected reactions that confuse and endanger others. In a recent analysis of California traffic incidents with AVs, in 57% of crashes, the AV was rear-ended by human drivers (3), with many of these crashes occurring because the AV behaved in an unexpected way that the human driver did not anticipate. For AVs to integrate onto roadways with human drivers, they must understand the intent of the human drivers and respond in a predictable and interpretable way.

While planning a left turn may be trivial for an AV on an empty roadway, it remains difficult in heavy traffic. For human drivers, these unprotected left turns often occur when an oncoming driver slows down to yield, an implicit signal to the other driver that it is safe to turn. An AV must also recognize these social cues of selfishness or cooperation, and failure to do so impacts the overall flow of the traffic network and even the safety of the traffic participants. AVs rely on explicit communication, state machines, or geometric reasoning about the driving interactions (4–8), neglecting social cues and driver personality. These approaches cannot handle complex interactions, resulting in conservative behavior and limiting autonomy solutions to simple road interactions. Additionally, humans cannot directly quantify and communicate their actions and decisions to autonomous agents. We use game theory to capture the dynamic interactions between agents, considering an agent's "best response" given the decisions of all other agents. Other approaches that use game-theoretic formulations model agents as selfish with homogeneous decision making (9–12). Instead, we extend the ability of AVs' reasoning by incorporating estimates of the other drivers' personality and driving style from social cues. This allows us to handle more complex navigation scenarios that rely on interactions, like multiple vehicles in an intersection. We present a mathematical formulation that combines control-theoretic approaches with models and metrics from the psychology literature, behavioral game theory, and machine learning.

Main Contributions. This article proposes a system to measure, quantify, and predict human behavior to better inform an

Significance

We present a framework that integrates social psychology tools into controller design for autonomous vehicles. Our key insight utilizes Social Value Orientation (SVO), quantifying an agent's degree of selfishness or altruism, which allows us to better predict driver behavior. We model interactions between human and autonomous agents with game theory and the principle of best response. Our unified algorithm estimates driver SVOs and incorporates their predicted trajectories into the autonomous vehicle's control while respecting safety constraints. We study common-yet-difficult traffic scenarios: highway merging and unprotected left turns. Incorporating SVO reduces error in predictions by 25%, validated on 92 human driving merges. Furthermore, we find that merging drivers are more competitive than nonmerging drivers.

Author contributions: W.S., A.P., J.A.-M., and D.R. designed research; W.S., A.P., J.A.-M., S.K., and D.R. performed research; W.S. and A.P. contributed new reagents/analytic tools; W.S., A.P., J.A.-M., S.K., and D.R. analyzed data; and W.S., A.P., J.A.-M., S.K., and D.R. wrote the paper.

Competing interest statement: W.S., A.P., J.A.-M., S.K., and D.R. are inventors on a provisional patent disclosure (filed by Massachusetts Institute of Technology) related to the social behavior for autonomous vehicles and uses thereof.

This article is a PNAS Direct Submission. P.A.M.V.L. is a guest editor invited by the Editorial Board.

[1]To whom correspondence may be addressed. Email: wilkos@mit.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1820676116/-/DCSupplemental.

First published November 22, 2019.

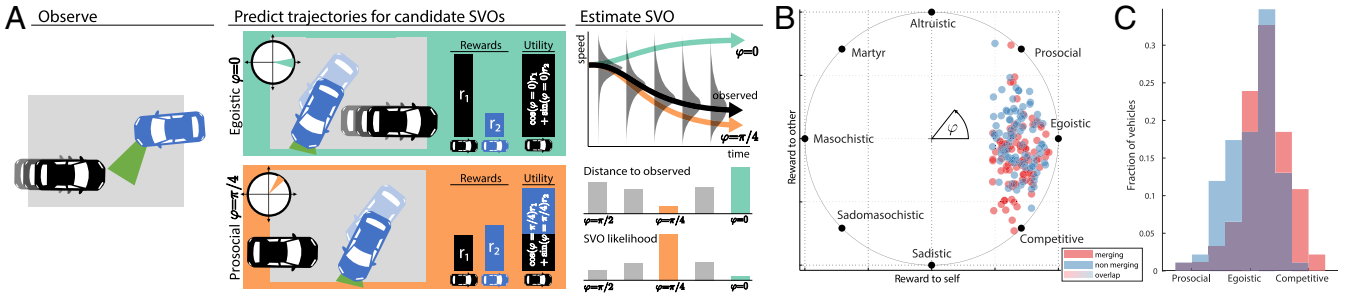

Fig. 1. (*A*) Knowing a driver's SVO helps predict their behavior. Here, the AV (blue) observes the trajectories of the other human driver (black). We can predict future motion of the black vehicle for candidate SVOs based on a utility-maximizing decision model (*Driving as a Game in Mixed Human–Robot Systems*). If the human driver is egoistic, they will not yield, and the AV must wait to turn. If the human driver is prosocial, they will yield, and the AV can safely turn. In both cases, the driver is utility-maximizing, but the utility function varies by SVO. An egoistic driver considers only its own reward in computing its utility. A prosocial driver weights its reward with the reward of the other car. The most likely SVO is the one that best matches a candidate trajectory to the actual observed trajectory (*Measuring and Estimating SVO Online*). The AV predicts future motion using the most likely SVO estimate. (*B*) SVO is represented as an angular preference φ that relates how individuals weight rewards in a social dilemma game. Here, we plot the estimated SVOs for drivers merging in the NGSIM dataset, explained in *Methods and Results*. (*C*) The distribution of mean SVO estimates during interactions. We find merging drivers (red) to be more competitive than nonmerging drivers (blue).

autonomous system. A game-theoretic formulation models driving as a series of social dilemmas to represent the dynamic interaction between drivers. We formulate a direct solution of the best-response game, allowing for fast, online predictions and planning, while integrating environmental and planning constraints to ensure safety. The game's reward functions are dynamic and dependent on the vehicles' states and the environment. Since we learn the reward functions from human driving data, we expect that our approach translates to other traffic scenarios and, broadly, human–robot interactions, where we can derive similar predictions trained on relevant data. Using Social Value Orientation (SVO), a common metric from psychology, we quantify human social preferences and their corresponding levels of cooperation. SVO measures how an individual weights their reward against the rewards of others, which translates into altruistic, prosocial, egoistic, or competitive preferences. We estimate the human drivers' SVOs from observed motion and set the AV's SVO based on the scenario.

The main contributions of this paper are as follows: modeling driving as a dynamic game and computing its Nash equilibrium; predicting human actions from expected utility maximization; integrating SVO preferences into the utility-maximizing framework; estimating SVO online from observed driving trajectories; simulations of emerging socially compliant autonomous driving behavior; and validation on Next Generation Simulation (NGSIM) (13) driving data, a human dataset of the US Highway 101.

Driving as a Game. We model driving as a noncooperative dynamic game (14), where the driving agents maximize their accumulated reward, or "payout," over time. At each point in time, the agent receives a reward, which may be defined by factors like delay, comfort, distance between cars, progress to goal, and other priorities of the driver. Fig. 1*A* illustrates an example of a driving game: an unprotected left turn. Here, the blue car must make a left turn across the path of the black car. Depending on how the interaction is resolved, the agents accrue different rewards for decisions such as comfortable braking, waiting for others to pass, or safety. In Fig. 1*A*, if each driver only maximizes their own reward, then the black vehicle would never brake for the blue vehicle making the unprotected left turn. However, we know that human drivers often brake for others in an act of altruism or cooperation. Similarly, in highway driving, we observe human drivers open gaps for merging vehicles. If all agents were to act in pure selfishness, the result would be increased congestion and, therefore, a decrease in the overall group's reward. We thus conclude that driving poses a sequence of social dilemmas.

Social Coordination. Social dilemmas involve a conflict between the agent's short-term self-interest and the group's longer-term collective interest. Social dilemmas occur in driving, where drivers must coordinate their actions for safe and efficient joint maneuvers. Other examples include resource depletion, low voter turnout, overpopulation, the prisoner's dilemma, or the public goods game. The autonomous control system proposed in this paper builds on social preferences of human drivers to predict outcomes of social dilemmas: whether individuals cooperate or defect, such as opening or closing a gap during a traffic merge. It allows us to better predict human behavior, thus offering a better basis for decision-making. It may also improve the efficiency of the group as a whole through emerging cooperation, for example, by reducing congestion.

Social Value Orientation. Behavioral and experimental economics shows that people have unique and individual social preferences, including interpersonal altruism, fairness, reciprocity, inequity aversion, and egalitarianism. Self-interested models, like the homo economicus (15), assume that agents maximize only their own reward in a game, which fails to account for nuances in real human behavior. In contrast, SVO indicates a person's preference of how to allocate rewards between themselves and another person. SVO can predict cooperative motives, negotiation strategies, and choice behavior (16–21). SVO preferences can be represented with a slider measure (22), a discrete-form triple dominance measure (23), or as an angle φ within a ring (24). We denote SVO in angular notation, shown in Fig. 1*B*.

Returning to Fig. 1*A*, SVO helps explain when the black car yields. Here, the black car considers both its reward and the reward of the blue car, weighted by SVO. As the angular preference increases from egoistic to prosocial, the weight of the other agent's reward increases, making it more likely that the black car will yield. Knowing a vehicle's SVO helps an AV better predict the actions of that vehicle and allows it to complete the turn if cooperation is expected. Without SVO, it would wait conservatively until all cars cleared the intersection.

An AV needs to estimate SVO, since humans cannot communicate this directly. Instead, humans observe and estimate SVO from actions and social cues (25). SVO preference distributions of individuals are largely individualistic ($\sim 40\%$) and prosocial ($\sim 50\%$) (22, 26–29), which emphasizes that an SVO-based model will be more accurate than a purely selfish model. We

estimate SVOs of other drivers by determining the SVO that best fits predicted trajectories to the actual driver trajectories. This technique enables the estimation and study of SVO distributions of agent populations directly from trajectory data, extending beyond driving. We plot the estimated SVOs for drivers merging in the NGSIM dataset in Fig. 1.

Socially Compliant Driving. Using SVO estimates of human drivers, we can design the control policy of the AV. We define *socially compliant driving* as behaving predictably to other human and autonomous agents during the sequence of driving social dilemmas. Achieving socially compliant driving in AVs is fundamental for the safety of passengers and surrounding vehicles, since behaving in a predictable manner enables humans to understand and appropriately respond to the AV's actions. To achieve socially compliant driving, the autonomous system must behave as human-like as possible, which requires an intrinsic understanding of human behavior, as well as the social expectations of the group. Human behavior may be imitated by learning human policies from data through *imitation learning* (30, 31). Our autonomous system design enables social compliance by learning human reward functions through *inverse reinforcement learning* (IRL) (32). The optimal control policy of the best-response game with learned rewards yields a human-imitating policy (9, 33, 34). Mathematically, the imitating policy is the expectation of human behavior based on past observed actions, capable of predicting and mimicking human trajectories. Combined with SVO, this enables an AV to behave as a human driver is expected to behave in traffic scenarios, such as acting more competitively during merges, and mirroring the utility-maximization strategies of humans with heterogeneous social preferences in social dilemmas (35).

When designing a cooperative AV, it may be desirable to assign the AV a prosocial SVO. Prosocials exhibit more fairness and considerateness compared to individualists (16) and engage in more volunteering, proenvironment, procommunity, and charitable efforts (17, 36–38). They also tend to minimize differences in outcomes between self and others (inequality aversion and egalitarianism) (18, 22). Additional findings suggest reciprocity in SVO and resulting cooperation (35, 39, 40).

To make the unprotected turn in Fig. 1*A*, the AV first observes the trajectory of the oncoming car, which can be done with onboard sensors. Using the reward (payoff) structure learned from data and our utility-maximizing behavior model, it generates candidate trajectories based on possible SVO values. The most likely SVO is the one that best matches a candidate trajectory to the actual observed trajectory. With this estimated SVO, the AV then generates future motion predictions and plans when to turn safely.

Estimating Driver Behavior with SVO

Our approach integrates SVO into a noncooperative dynamic game, and we model the agents as making utility-maximizing decisions, with the optimization framework presented in *Driving as a Game in Mixed Human–Robot Systems*. To integrate SVO into our game-theoretic formulation, we define a utility function $g(\cdot)$ that combines the rewards of the ego agent with other agents, weighted by the ego agent's SVO angular preference φ. For a two-agent game,

$$g_1 = \cos(\varphi_1) r_1(\cdot) + \sin(\varphi_1) r_2(\cdot), \quad [1]$$

where r_1 and r_2 are the "reward to self" and "reward to other," respectively, and φ_1 is the ego agent's SVO. We see that the orientation of φ_1 will weight the reward r_1 against r_2 based on the ego agent's actions. The following definitions of social preferences (22, 24) are based on these weights:

1) Altruistic: Altruistic agents maximize the other party's reward, without consideration of their own outcome, with $\varphi \approx \frac{\pi}{2}$.
2) Prosocial: Prosocial agents behave with the intention of benefiting a group as a whole, with $\varphi \approx \frac{\pi}{4}$. This is usually defined by maximizing the joint reward.
3) Individualistic/egoistic: Individualistic agents maximize their own outcome, without concern of the reward of other agents, with $\varphi \approx 0$. The term egoistic is also used.
4) Competitive: Competitive agents maximize their relative gain over others, i.e., $\varphi \approx -\frac{\pi}{4}$.

We limit our definitions to rational social preferences, with more in refs. 22 and 24. While our definitions give specific values of SVO preferences for clarity, we also note that SVO exists on a continuum. For example, values in the range $0 < \varphi < \frac{\pi}{2}$ all exhibit a certain degree of altruism. We denote *cooperative actions* as actions that improve the outcome for all agents. For example, two egoistic agents may cooperate if both benefit in the outcome. Prosocials make cooperative choices, as their utility-maximizing policy also values a positive outcome of others. These cooperative choices improve the efficiency of the interaction and create collective value.

Measuring and Estimating SVO Online. Given that other drivers maximize utility, we can predict their trajectories from observations and an estimate of their SVO. The choice of SVO changes the predicted trajectories. In Fig. 1*A*, a prosocial SVO generates a braking trajectory prediction, while an egoistic SVO generates a nonbraking trajectory. We compute the likelihood of candidate SVOs from evaluating the Gaussian kernel on the distance between predicted and actual trajectories. We also consider a maximum-entropy model, which builds a likelihood function based on the distance of the observed trajectory to optimality given a candidate SVO (see *SI Appendix*, section S3 for derivations). We utilize these methods to estimate SVO from human driver trajectories in *Methods and Results*.

Benefit of SVO. We improve predictions of interactions by estimating the SVO of other drivers online. Incorporating SVO into the model increases social compliance of vehicles in the system, by improving predictability and blending in better. For the AVs, SVO adds the capability of nuanced cooperation with only a single variable. The AV's SVO can be specified as user input, or change dynamically according to the driving scenario, such as becoming more competitive during merging.

Driving as a Game in Mixed Human–Robot Systems

To create a socially compliant autonomous system, our autonomous agents must determine their control strategies based on the decisions of the human and other agents. This section details how we incorporate a human decision-making model into an optimization framework; see *SI Appendix*, section S2 for more detail. We formulate the utility-maximizing optimization problem as a multiagent dynamic game and then derive the Nash equilibrium to solve for a socially compliant control policy.

Consider a system of m human drivers and autonomous agents, with states such as position, heading, and speed, at time k denoted $\mathbf{x}_i^k \in \mathcal{X}$, where $i = \{1, \ldots, m\}$ and $\mathcal{X} \in \mathbb{R}^n$ is the set of all possible states. We denote $\mathbf{u}_i^k \in \mathcal{U}$ as the control input, such as acceleration and steering angle, of agent i and $\varphi_i \in \Phi$ as SVO preference, where $\mathcal{U} \in \mathbb{R}^n$ is the set of all possible control inputs and Φ is the set of possible SVO preferences. For brevity, we write the state of all agents in the system as $\mathbf{x} = [\mathbf{x}_1^\intercal, \ldots, \mathbf{x}_m^\intercal]^\intercal$, all control inputs as $\mathbf{u} = [\mathbf{u}_1^\intercal, \ldots, \mathbf{u}_m^\intercal]^\intercal$. The states evolve according

Table 1. Trajectory prediction error

		Multiagent game theoretic		
Prediction	Baseline	1	2	3
SVO	—	Egoistic	Static best	Estimated
MSE position	1.0	0.947	0.821	0.753

Relative position mean squared error (MSE) between predicted and actual trajectories, compared to a single-agent baseline. Our multiagent game-theoretic model reduces error, and the dynamic, estimated SVO performs best.

to dynamics $\mathcal{F}_i(\mathbf{x}_i^k, \mathbf{u}_i^k)$ subject to constraints $c_i(\cdot) \leq 0$ with the discrete-time transition function

$$\mathbf{x}^{k+1} = \mathcal{F}(\mathbf{x}^k, \mathbf{u}^k) = \left[\mathcal{F}_1(\mathbf{x}_1^k, \mathbf{u}_1^k)^\intercal, \ldots, \mathcal{F}_m(\mathbf{x}_m^k, \mathbf{u}_m^k)^\intercal\right]^\intercal. \quad [2]$$

The notation $\mathbf{x}_{\neg i}$ refers to the set of agents excluding agent i. For example, we can write the state vector $\mathbf{x} = [x_1^\intercal \mid x_{\neg 1}^\intercal]^\intercal$, with $\mathbf{x}_{\neg 1} = [\mathbf{x}_2^\intercal, \ldots, \mathbf{x}_m^\intercal]^\intercal$. The agents calculate their individual control policies $\mathbf{u}_i$ by solving a general discrete-time constrained optimization over N time steps and time horizon $\tau = \sum_{k=1}^{N} \Delta t$. The set of states over the horizon is denoted as $\mathbf{x}^{0:N}$, and the set of inputs is $\mathbf{u}^{0:N-1}$. To calculate the control policy, we formulate a utility function for each agent and then find the utility-maximizing control actions. The utility function is defined as a combination of reward functions $r_i(\cdot)$, as described in Eq. **1**, and calculated from weighted features of the current state, controls, the environment, and social preference φ_i. At a given time k, each agent i's utility function is given by $g_i\left(\mathbf{x}^k, \mathbf{u}^k, \varphi_i\right)$, and $g_i^N\left(\mathbf{x}^N, \varphi_i\right)$. The utility over the time horizon τ is denoted $G_i(\cdot)$, written

$$G_i(\mathbf{x}^0, \mathbf{u}, \varphi_i) = \sum_{k=0}^{N-1} g_i\left(\mathbf{x}^k, \mathbf{u}^k, \varphi_i\right) + g_i^N\left(\mathbf{x}^N, \varphi_i\right). \quad [3]$$

In this paper, we learn the reward functions $r_i(\cdot)$ from the NGSIM driving data to approximate real human behavior; see *SI Appendix*, section S3 for more details on this approach.

Human Decision-Making Model. From psychology literature, we find that people are heterogeneous in their evaluation of joint rewards (18), and we can model preferences for others using utility functions that weight rewards (39–41). Murphy and Ackermann (35) model human decision making as expected utility maximizing under individual preferences. Based on these findings from behavioral decision theory, we model human agents in our system as agents that make utility-maximizing decisions. Other robotics literature (9, 33, 34) supports this case. Translating this decision making into an optimization framework for socially compliant behavior, we write the utility-maximizing policy

$$\mathbf{u}_i^*\left(\mathbf{x}^0, \varphi_i\right) = \arg\max_{\mathbf{u}_i} \; G_i\left(\mathbf{x}^0, \mathbf{u}_i, \mathbf{u}_{\neg i}, \varphi_i\right). \quad [4]$$

The solution $\mathbf{u}_i^*$ to Eq. **4** also corresponds to the actions maximizing the likelihood under the maximum entropy model

$$P(\mathbf{u}_i | \mathbf{x}^0, \mathbf{u}_{\neg i}, \varphi_i) \propto \exp\left(G_i(\mathbf{x}^0, \mathbf{u}_i, \mathbf{u}_{\neg i}, \varphi_i)\right), \quad [5]$$

used to learn our rewards by IRL (32, 42). Under this model, the probability of actions $\mathbf{u}$ is proportional to the exponential of the utility encountered along the trajectory. Hence, utility maximization yields actions most likely imitating human driver behavior, which is important for social compliance.

Although the human driver does not explicitly calculate $\mathbf{u}$, we assume our model and formulation of $\mathbf{u}$ captures the

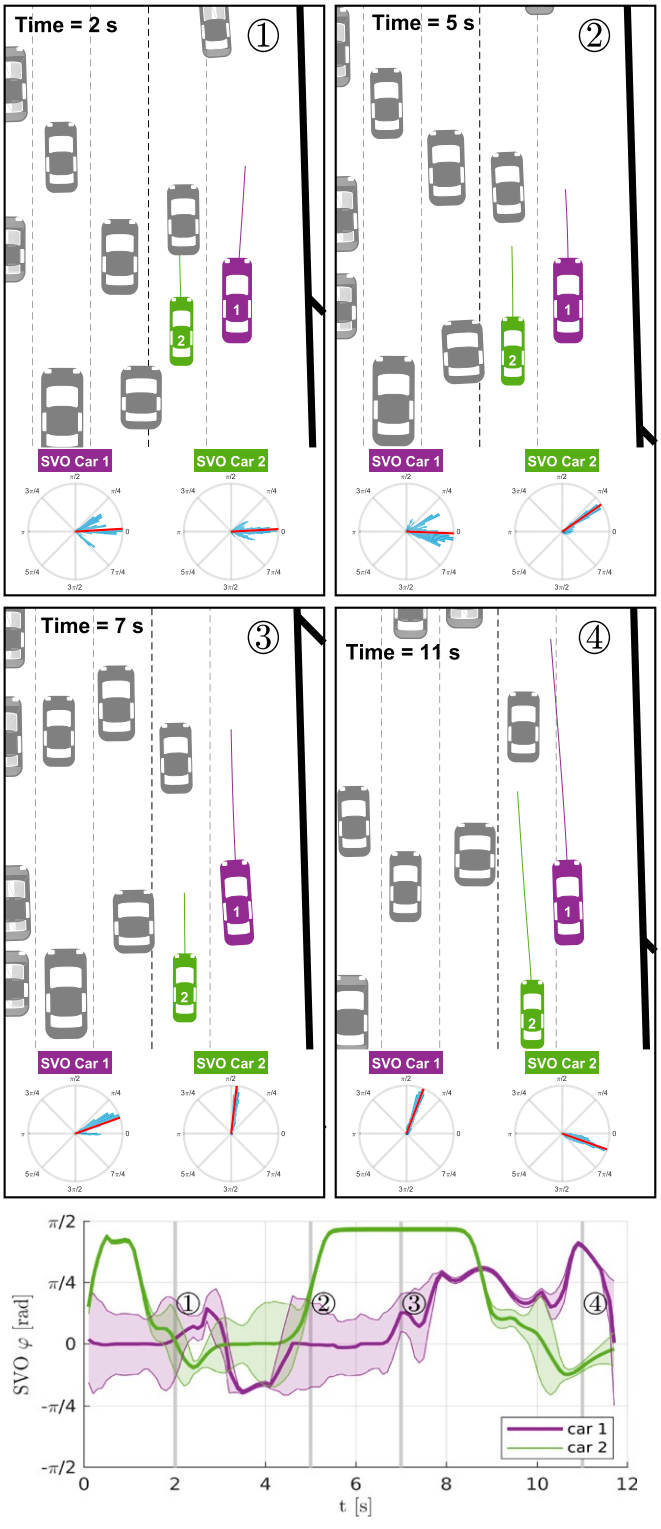

Fig. 2. (*Upper*) Snapshot of NGSIM dataset with $n = 2$ active cars (purple and green) and $n = 50$ obstacle cars (gray). Here, car 1 (purple) is attempting a merge and must interact with car 2 (green). The solid lines indicate the predicted trajectory from our algorithm. For SVO estimates at each frame, the blue represents the distribution, while the red line indicates our estimate. (*Lower*) The solid line indicates SVO estimate over time, with the shaded region representing the confidence bounds. Initially, car 2 does not cooperate with car 1 and does not allow it to merge. After a few seconds, car 2 becomes more prosocial, which corresponds to it "dropping back" and allowing the first car to merge.

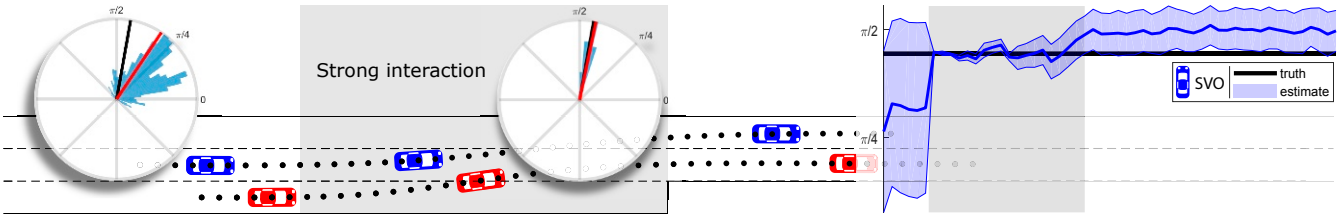

Fig. 3. (*Left*) Estimated distribution of SVO preference of blue car shown as polar histograms in SVO circles for premerge and during merge. (*Right*) The mean estimate is shown as red and the ground truth (80°, altruistic) in black. SVO estimates with 1-σ uncertainty bounds are shown on the right. Area of strong interaction corresponds to gray area on both sides.

decision-making process of the human driver based on their observations, control actions, and underlying reward function $r_i(\cdot)$ of the environment. Later, we validate on the NGSIM dataset that our learned model successfully predicts the actual trajectories driven by the human drivers.

Game-Theoretic Autonomous Control Policy with SVO. To design the control policy for the AV, note that Eq. **4** formulated for all m agents simultaneously defines a dynamic game (14). Given SVO estimates for all agents and a set of constraints on the system, we solve for the optimal control policy of a vehicle, u_i^*, assuming the other agents in the system also choose an optimal policy, $u_{\neg i}^*$. For an intuition on how these dynamic games work, we first start with a Stackelberg game. An example traffic scenario that can be modeled as a Stackelberg game is cars arriving at a four-way stop, where they must traverse the intersection based on the first arrival. In the traditional two-agent Stackelberg game (43), the leader $(i=1)$ makes its choice of policy, $\mathbf{u}_1$, and the follower $(i=2)$ maximizes their control given the leader policy, $\mathbf{u}_2^*(\mathbf{u}_1)$. See *SI Appendix*, section S2 for details on the general procedure of a multi-agent Stackelberg game. While the Stackelberg game can model some intersections, in many traffic scenarios, it is unclear who should be the leader and the follower, thus necessitating a more symmetric and simultaneous choice game, which is the approach we use in this paper. In the two-agent case, the follower chooses $\mathbf{u}_2(\mathbf{u}_1)$, but the leader readjusts based on the follower, or $\mathbf{u}_1(\mathbf{u}_2)$. This back and forth creates more levels of tacit negotiation and best response, such that $\mathbf{u}_2(\mathbf{u}_1(\mathbf{u}_2(\mathbf{u}_1(\ldots))))$. This strategy removes the leader–follower dynamics, as well as any asymmetric indirect control, yielding a simultaneous choice game.

Nash Equilibrium. The iterative process of exchanging and optimizing policies is also called iterative best response, a numerical method to compute a Nash equilibrium of the game defined by Eq. **4**. A limitation is its iterative nature; optimizing may take an unacceptable amount of steps. To make solving for the Nash equilibrium computationally tractable, we reformulate the m interdependent optimization problems as a local single-level optimization using the Karush–Kuhn–Tucker conditions (14, 44). We solve the locally equivalent formulation, including all constraints, with state-of-the-art nonlinear optimizers. This preserves all safety constraints in the optimization, critical for guaranteeing safe operation, and performance Algorithm 1 provides an overview of our method, with more details in *SI Appendix*, section S3.

The Nash equilibrium yields a control law for the AV $\mathbf{u}_i^*$ as well as predicted actions $\mathbf{u}_{\neg i}^*$ for all other $m-1$ agents N time steps into the future. Based on learned reward functions and the maximum entropy model, Eq. **5**, $\mathbf{u}_{\neg i}^*$ are also maximum likelihood predictions. The Nash equilibrium is the predicted outcome of the driving social dilemma and mimics the negotiation process between agents.

Methods and Results

We implement our socially compliant driving algorithm in two ways: first to predict human driver behavior in highway merges, then in simulations of autonomous merging and turning scenarios. This section highlights illustrative examples of our results, with expanded analysis included in *SI Appendix*, section S7. We evaluate human driver predictions on the NGSIM dataset and examine highway on-ramp merges into congestion. We analyze a total of 92 unique merges from the dataset and discuss key results on a representative example. Incorporating SVO reduces errors in trajectory predictions of human drivers by up to 25%. For the AV simulations, we replicate this merging scenario and also present an unprotected left turn. Our simulations demonstrate how using SVO preferences assists the AV in choosing safe actions, adding nuanced behavior and cooperation with a single parameter.

Predicting Human Driving Behavior. To validate our algorithm, we test its ability to predict human trajectories on highway on-ramp merges in the NGSIM dataset. We implement a noninteractive baseline algorithm, where each agent computes their optimal policy while modeling other agents as lane-keeping dynamic obstacles. Using the dataset and trajectory history, we compare the baseline prediction's performance to the multiagent game-theoretic models with: 1) static egoistic SVO, equal to

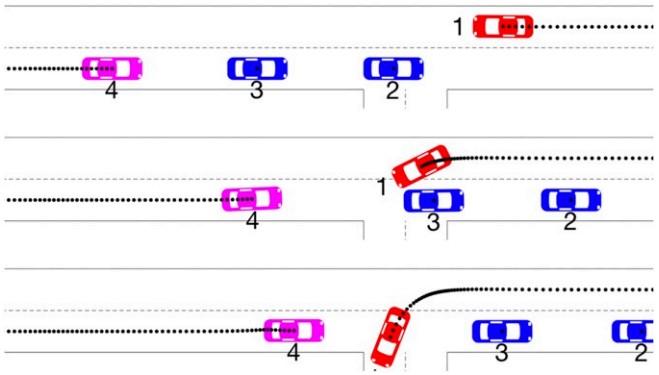

Fig. 4. Unprotected left turn of an AV (red; $i=1$) with oncoming traffic. As the AV approaches the intersection, two egoistic cars (blue; $i=2, 3$) continue and do not yield. A third altruistic car (magenta; $i=4$) yields by slowing down, allowing the AV to complete the turn in the gap.

Algorithm 1: Socially Compliant Autonomous Driving:

1: $\mathbf{x}^0 \leftarrow$ Update state observations of all agents
2: $\varphi_{\neg 1} \leftarrow$ Update SVO estimation of all agents
3: $\varphi_1 \leftarrow$ Choose AV SVO
4: $\mathbf{u}^* \leftarrow$ Plan and predict for all agents Eq. **4** (multiagent Nash equilibrium)
5: Execute AV's optimal control $\mathbf{u}_1^*$

neglecting the SVO model; 2) best static SVO; and 3) estimated dynamic SVOs. The best static SVO corresponds to the best SVO estimate when holding it constant throughout the interaction. For different interactions, this may yield a different static SVO. Table 1 examines the relative position error between the true vehicle trajectory and our predictions. We find that incorporating the multiagent game-theoretic framework, but remaining egoistic, alone improves performance by 5%. Highlighting the importance of SVO, we see an 18% improvement over the baseline with static SVO and 25% with estimated dynamic SVO.

Fig. 2 shows a two-agent merge with car 1 (purple) merging into the lane with car 2 (green). We model the other cars in the dataset as obstacles for the planner. For a dynamic SVO prediction, we estimate SVO online from observed trajectories of the vehicles, then leverage SVO in predicting the trajectory. Fig. 2 shows SVO predictions and confidence bounds for both cars through the merge. Our SVO estimates help explain the interactions occurring: At $t = 2$, the first car's SVO is egoistic while attempting to merge, but the second car is also egoistic and does not provide a sufficient gap to merge. At $t = 5$, the second car drops back and increases the gap for merging, corresponding to a more prosocial estimated SVO. Once the first car has merged, the second car closes the gap, returning to an egoistic SVO.

The capability of estimating SVOs of humans by observing their motions allows us to investigate how SVO distributions in natural populations differ. Separating merging and nonmerging vehicles in the dataset, we find that merging cars are more likely to be competitive than nonmerging cars, as shown in the histogram of Fig. 1*C*. This observation also withstands hypothesis testing with statistical significance ($P < 0.002$), further discussed in *SI Appendix*, section S7.

Autonomous Merging with SVO. Employing the estimation techniques described in *SI Appendix*, section S3, we are able to measure SVO preference of another agent in a simulated highway-merging scenario. Fig. 3 shows the AV's (red) SVO estimates of another vehicle (blue) over time. At first, the vehicles have little interaction, and the observations of the driver's SVO remain ambiguous, such that the estimate is inaccurate with high variance. As the AV approaches the end of its lane, both vehicles begin to interact, indicated in gray in the figure. During this time, the SVO estimate quickly converges to the true value, with high confidence. After the merge, the vehicles no longer interact, the variance of the SVO estimate increases, and the estimate drifts away from the true value. Note that estimating the characteristics of an interaction (e.g., SVO) is only possible if the interaction between agents is impactful; see *SI Appendix*, section S4 for a Hessian-based analysis.

Unprotected Left Turns. In this scenario, the AV must make an unprotected left turn against numerous cars traveling in the oncoming direction. If the AV were in light traffic, it could be feasible for it to wait for all other oncoming cars to pass. However, in congested traffic, the intersection might never fully clear. Instead, the AV must predict when an oncoming car will yield, allowing the vehicle to safely make the turn. Fig. 4 shows our simulation, where the AV (red; $i = 1$) attempts to turn across traffic. Two egoistic cars (blue; $i = 2, 3$) approach the intersection and do not yield for the AV, as predicted. An altruistic third car (magenta; $i = 4$) yields for the AV by slowing down, such that the gap between itself and the other blue car increases. With this increased gap, the AV is able to safely make the turn, and the magenta car continues forward.

Conclusions

We propose the use of SVO to measure, quantify, and predict the behavior of human drivers. We model the interactions between drivers as a dynamic game and present a computationally tractable way of finding its Nash equilibrium. Using SVO as our key factor in predicting human behavior, we present two likelihood functions to estimate the SVO of other drivers from observed trajectories. We validate our findings in simulation and on the NGSIM dataset, incorporating the human behavior into the AV planner, resulting in intelligent, socially aware maneuvers. We find that the multiagent Nash equilibrium, SVO, as well as its estimation improve predictions and prove essential assets for interactive driving. Our unified algorithm improves on human driver trajectory prediction by 25% over baseline models. For highway merges in the NGSIM data, we also find that the human drivers merging into traffic are consistently more competitive than the drivers yielding to the merging car. These insights can better inform AVs that currently struggle to make these maneuvers. The ability to estimate SVO distributions directly from observed motion instead of in laboratory conditions will prove impactful beyond autonomous driving. Overall, robotic and artificial intelligence applications where an autonomous system acts among humans will benefit from integrating SVO in their prediction and decision-making algorithms.

Data Availability Statement. The NGSIM data are available at ref. 13.

ACKNOWLEDGMENTS. We thank Brandon Araki for helpful insight. This research was supported in part by the Netherlands Organisation for Scientific Research, Applied and Engineering Sciences, and Toyota Research Institute (TRI). This article solely reflects the opinions and conclusions of its authors and not TRI, Toyota, or any other entity. We thank them for this support.

1. National Highway Traffic Safety Administration, "*Traffic Safety Facts 2015*" (Tech. Rep. DOT HS 812 318, National Highway Traffic Safety Administration, Washington, DC, 2015), p. 101.
2. A. Efrati, Waymo's big ambitions slowed by tech trouble. *The Information* (2018). https://www.theinformation.com/articles/waymos-big-ambitions-slowed-by-tech-trouble.
3. J. Stewart, Why people keep rear-ending self-driving cars. *Wired* (2018) https://www.wired.com/story/self-driving-car-crashes-rear-endings-why-charts-statistics/.
4. C. Urmson *et al.*, Autonomous driving in urban environments: Boss and the urban challenge. *J. Field Robot.* **25**, 425–466 (2008).
5. J. Leonard *et al.*, A perception-driven autonomous urban vehicle. *J. Field Robot.* **25**, 727–774 (2008).
6. M. Düring, P. Pascheka, "Cooperative decentralized decision making for conflict resolution among autonomous agents" in *2014 IEEE International Symposium on Innovations in Intelligent Systems and Applications (INISTA) Proceedings* (IEEE, Piscataway, NJ, 2014), pp. 154–161.
7. W. Schwarting, P. Pascheka, "Recursive conflict resolution for cooperative motion planning in dynamic highway traffic" in *17th International IEEE Conference On Intelligent Transportation Systems (ITSC)* (IEEE, Piscataway, NJ, 2014), pp. 1039–1044.
8. A. Pierson, L. C. Figueiredo, L. C. Pimenta, M. Schwager, Adapting to sensing and actuation variations in multi-robot coverage. *Int. J. Robot. Res.* **36**, 337–354 (2017).
9. D. Sadigh, S. Sastry, S. A. Seshia, A. D. Dragan, "Planning for autonomous cars that leverage effects on human actions" in *Proceedings of Robotics: Science and Systems*, D. Hsu, N. Amato, S. Berman, S. Jacobs, Eds. http://www.roboticsproceedings.org/rss12/p29.html. Accessed 15 November 2019.
10. R. Spica *et al.*, "A real-time game theoretic planner for autonomous two-player drone racing" in, *Proceedings of Robotics: Science and Systems*. http://www.roboticsproceedings.org/rss14/p40.pdf. Accessed 15 November 2019.
11. G. Williams, B. Goldfain, P. Drews, J. M. Rehg, E. A. Theodorou, "Best response model predictive control for agile interactions between autonomous ground vehicles" in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, K. Lynch, *et al.*, Eds. (IEEE, Piscataway, NJ, 2018), pp. 2403–2410.
12. A. Liniger, J. Lygeros, A noncooperative game approach to autonomous racing. *IEEE Trans. Control Syst. Technol.*, 8643396 (2019).
13. Federal Highway Administration, US Department of Transportation, "NGSIM: Next generation simulation" (Federal Highway Administration, Washington, DC, 2017).
14. T. Basar, G. J. Olsder, *Dynamic Noncooperative Game Theory* (Classics in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia) vol. 23 (1999).
15. C. F. Camerer, E. Fehr, When does "economic man" dominate social behavior? *Science* **311**, 47–52.
16. C. K. De Dreu, P. A. Van Lange, The impact of social value orientations on negotiator cognition and behavior. *Personal. Soc. Psychol. Bull.* **21**, 1178–1188 (1995)

17. C. G. McClintock, S. T. Allison, Social value orientation and helping behavior. *J. Appl. Soc. Psychol.* **19**, 353–362 (1989).
18. P. A. Van Lange, The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *J. Personal. Soc. Psychol.* **77**, 337–349 (1999)
19. R. O. Murphy, K. A. Ackermann, Social value orientation: Theoretical and measurement issues in the study of social preferences. *Personal. Soc. Psychol. Rev.* **18**, 13–41 (2014)
20. J. L. Pletzer *et al.*, Social value orientation, expectations, and cooperation in social dilemmas: A meta-analysis. *Eur. J. Personal.* **32**, 62–83 (2018)
21. K. A. Ackermann, R. O. Murphy, Explaining cooperative behavior in public goods games: How preferences and beliefs affect contribution levels. *Games* **10**, 15 (2019).
22. R. O. Murphy, K. A. Ackermann, M. Handgraaf, Measuring social value orientation. *Judgment Decis. Making* **6**, 771–781 (2011).
23. P. A. Van Lange, E. De Bruin, W. Otten, J. A. Joireman, Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *J. Personal. Soc. Psychol.* **73**, 733–746 (1997).
24. W. B. G. Liebrand, C. G. McClintock, The ring measure of social values: A computerized procedure for assessing individual differences in information processing and social value orientation. *Eur. J. Personal.* **2**, 217–230 (1988).
25. G. P. Shelley, M. Page, P. Rives, E. Yeagley, D. M. Kuhlman, "Nonverbal communication and detection of individual differences in social value orientation" in *Social Decision Making: Social Dilemmas, Social Values, and Ethical Judgments*, R. M. Kramer, A. E. Tenbrunsel, M. H. Bazerman, Eds. (Organization and Management Series, Psychology Press, New York, 2009), pp. 147–169.
26. D. Balliet, C. Parks, J. Joireman, Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Process. Intergr. Relat.* **12**, 533–547. (2009).
27. A. Garapin, L. Muller, B. Rahali, Does trust mean giving and not risking? Experimental evidence from the trust game. *Rev. Écon. Polit.* **125**, 701–716 (2015).
28. J. P. Carpenter, Is fairness used instrumentally? Evidence from sequential bargaining. *J. Econ. Psychol.* **24**, 467–489 (2003).
29. S. Fiedler, A. Glöckner, A. Nicklisch, S. Dickert, Social value orientation and information search in social dilemmas: An eye-tracking analysis. *Organ. Behav. Hum. Decis. Process.* **120**, 272–284 (2013).
30. S. Ross, G. Gordon, D. Bagnell, "A reduction of imitation learning and structured prediction to no-regret online learning" in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, G. Gordon, D. Dunson, M. Dudík, Eds. (Proceedings of Machine Learning Research, Fort Lauderdale, FL, 2011), vol. 15, pp. 627–635.
31. J. Ho, S. Ermon, Generative adversarial imitation learning in *Advances in Neural Information Processing Systems 29*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, R. Garnett, Eds. (Neural Information Processing Systems Foundation, 2016), pp. 4565–4573.
32. B. D. Ziebart, A. L. Maas, J. A. Bagnell, A. K. Dey, "Maximum entropy inverse reinforcement learning" in *Proceedings of the 23rd AAAI Conference on Artificial Intelligence*, A. Cohn, Ed. (Association for the Advancement of Artificial Intelligence, Palo Alto, CA, 2008), vol. 8, pp. 1433–1438.
33. H. Kretzschmar, M. Spies, C. Sprunk, W. Burgard, Socially compliant mobile robot navigation via inverse reinforcement learning. *Int. J. Robot. Res.* **35**, 1289–1307 (2016).
34. B. D. Ziebart *et al.*, "Planning-based prediction for pedestrians" in *2009 IEEE/RSJ International Conference On Intelligent Robots and Systems* (IEEE, Piscataway, NJ, 2009), pp. 3931–3936.
35. R. O. Murphy, K. A. Ackermann, Social preferences, positive expectations, and trust based cooperation. *J. Math. Psychol.* **67**, 45–50 (2015).
36. T. Gärling, S. Fujii, A. Gärling, C. Jakobsson, Moderating effects of social value orientation on determinants of proenvironmental behavior intention. *J. Environ. Psychol.* **23**, 1–9 (2003).
37. P. A. Van Lange, R. Bekkers, T. N. Schuyt, M. V. Vugt, From games to giving: Social value orientation predicts donations to noble causes. *Basic Appl. Soc. Psychol.* **29**, 375–384 (2007).
38. J. D'Attoma, C. Volintiru, A. Malezieux, Gender, social value orientation, and tax compliance (CESifo Working Paper 7372, CESifo, Munich, 2018).
39. K. A Ackermann, J. Fleiß, R. O. Murphy, Reciprocity as an individual difference. *J. Confl. Resolut.* **60**, 340–367 (2016).
40. F. Moisan, R. ten Brincke, R. Murphy, C. Gonzalez, Not all prisoner's dilemma games are equal: Incentives, social preferences, and cooperation. *Decision* **5**, 306–322 (2017).
41. J. Andreoni, J. Miller, Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica* **70**, 737–753 (2002).
42. S. Levine, K. Vladlen, Continuous inverse optimal control with locally optimal examples" in *Proceedings of the 29th International Conference on Machine Learning*, J. Langford, J. Pineau, Eds. (Omnipress, 2012), pp. 475–482.
43. H .Von Stackelberg, *Market Structure and Equilibrium* (Springer Science & Business Media, New York, 2010).
44. M. Pilecka *et al.*, Combined reformulation of bilevel programming problems. *Schedae Inf.* **21**, 65–79 (2012).