Modeling Supervisor Safe Sets for Improving Collaboration in Human-Robot Teams

David L. McPherson*, Dexter R.R. Scobee*, Joseph Menke, Allen Y. Yang, S. Shankar Sastry

***Abstract*— When a human supervisor controls a team of robots, their attention is divided and cognitive resources are at a premium. We aim to optimize the distribution of these resources and the flow of attention. We propose the model of an idealized noisy supervisor to describe human behavior. Such a supervisor possess an internal model of the the robots' dynamics. We use reachability theory to generate safe sets that represent supervisor behavior. The idealized supervisor will intervene to assist a robot if they perceive that it has left the supervisor's modeled safe set, regardless of whether or not the robot remains within a ground-truth safe set. False positives, where the human judges the robot to be in danger when the robot is actually safe, needlessly consume supervisor effort. We present a learning algorithm to estimate the supervisor's internal safe set, and a control algorithm for robot team members to respect the supervisor's safe set. We demonstrate that robot teams that behave according to this model will reduce the occurrence of false positives for our idealized supervisor model. Furthermore, we validate our approach with a user study that demonstrates a significant ($p = 0.0328$) reduction in false positives for our method compared to the baseline safety controller.**

I. Background And Introduction

As automation becomes more pervasive throughout society, humans will increasingly find themselves interacting with autonomous and semi-autonomous systems. These interactions have the potential to multiply the productivity of humans workers, since it will become possible for a single human to supervise the behavior of multiple robotic agents. For example, in the case of a single human driver managing a fleet of multiple delivery robots, the robots would be able to autonomously navigate the majority of their journeys, but the driver would be able to take control for the "last mile," guiding the robots to precisely deposit packages in environments where autonomous navigation may not be reliable. Human experts regularly serve as failsafe supervisors on factory assembly floors staffed with robotic arms [13]. Air traffic controllers soon will have to manage completely autonomous drones flying through their airspace alongside already traditional mixed-autonomy planes and their auto-pilots [19].

While a human may be able to successfully exert direct control over a single robot, it becomes intractable for a human to directly control teams of robots (in fact, humans often benefit from automated assistance when controlling even a single robot, as discussed in the literature on assistive teleoperation [7], [15]). In order to manage the increased complexity of multi-robot teams, the human must be able to rely on increased autonomy from the robots, freeing the human to focus their attention only on those areas where they are most needed. Our goal is to model what grabs the supervisor's attention in order to modify robot behavior to reduce the occurrence of distractions.

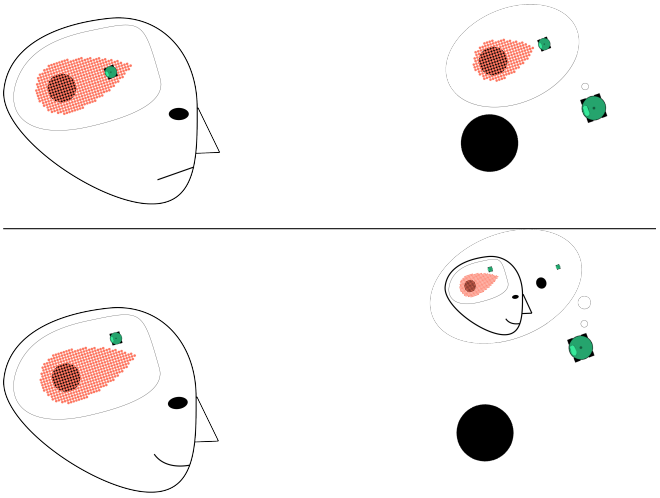

Fig. 1. Top: if a robot's behavior does not take into account a human supervisor's notion of safety, the misaligned expectations can degrade team performance. Bottom: When a robot acts according to a human supervisor's expectations, the supervisor can more easily predict the robot's behavior.

This project is inspired by work like Bajscy et al [3] and Jain et al [14] that learn from supervisor interventions in a "coactive" learning framework. These works apply Learning from Demonstration techniques to the more challenging domain where the given data is just a correction from a trajectory rather than a full trajectory. We favor a model-based approach that interprets the signal as the result of an optimization problem. Inverse Reinforcement Learning [1], [20] fits this inverse optimization framework to the dynamic decision-making context of Inverse Optimal Control as conceived of by Kalman [16]. Our work applies this inverse optimization framework to learn from the supervisor's decisions to intervene.

Results in cognitive science suggest that humans observing physical scenes can be modeled as performing a noisy "mental simulation" to predict trajectories [4], [18]. We posit that human supervisors utilize this same cognitive dynamic simulation to predict robot safety and intervene accordingly.

*The first two authors contributed equally

This work is supported by the Office of Naval Research under the Embedded Humans MURI (N00014-13-1-0341).

All authors are with the Department of Electrical Engineering and Computer Science, University of California, Berkeley {david.mcpherson, dscobee, joemenke, yang, sastry}@eecs.berkeley.edu

Specifically, we theorize that the intervention behavior is driven by an internal "safe set" which we can attempt to reconstruct from by observing supervisor interventions.

Safe sets are predicated upon the Formal Methods notion of "Viability". A set of states is "viable" if for every state in the set there exists a dynamic trajectory that stays within the set for all time. Reachability analysis calculates the largest viable set that doesn't include any undesirable state configurations (e.g. collisions with obstacles, power overloads, etc). Since the set is viable, it is possible to guarantee that the dynamic system will always stay within the set and therefore stay safely away from the undesirable states. For this reason, viability kernels are often refered to as "safe sets". Reachability can be used for robust path planning in dynamically changing environments [9] or working around multiple dynamic agents [5], and recent results have leveraged the technique to bound tracking error in order to generate dynamically feasible paths using simple planning algorithms [11].

Hoffman et al. used the safety guarantees of reachability analysis to engineer a multi-drone team that could automatically avoid collisions [12]. Similarly, Gillula could guarantee safety for learning algorithms by constraining their explorations to stay within the safe set [10]. Extending this, Akametalu and Tomlin [2] were able to guarantee safety while simultaneously learning and expanding the safe set. All of these controllers supervise otherwise un-guaranteed systems and intervene to maintain safety whenever the system threatens to leave the viable safe set. In this paper, we explore how this intervention behavior is similar to human supervision, and apply this to representing human safety concerns as safe sets in the state space.

II. Supervisor Safe Set Control

Based on the success of cognitive dynamical models for explaining humans' understanding of physical systems, we posit that human operators may have some notion of reachable sets which they employ to predict collisions or avoid obstacles. We propose a noisy idealized model to describe the behavior of the human supervisor of a robotic team, and we develop a framework for estimating the human supervisor's mental model of a dynamical system based on observing their interactions with the team. We then propose a control framework that capitalizes on this learned information to improve collaboration in such human-robot teams.

A. Preliminaries: Reachability for Safety

Consider a dynamical system with bounded input u and bounded disturbance d, given by

$$\begin{gathered}\dot{x} = f(x,u,d),\\ x\in\mathbb{R}^n,\quad u\in\mathcal{U}\subset\mathbb{R}^{n_u},\quad d\in\mathcal{D}\subset\mathbb{R}^{n_d},\end{gathered} \tag{1}$$

where $\mathcal{U}$ and $\mathcal{D}$ are compact. We let $\mathscr{U}$ and $\mathscr{D}$ denote the sets of measurable functions $\mathbf{u}:[0,\infty)\to\mathcal{U}$ and $\mathbf{d}:[0,\infty)\to\mathcal{D}$, respectively, which represent possible time histories for the system input and disturbance. Given a choice of input and disturbance signals, there exists a unique continuous trajectory $\xi:[0,\infty)\to\mathbb{R}^n$ from any initial state x which solves

$$\begin{aligned}&\dot{\xi}(t) = f(\xi(t),\mathbf{u}(t),\mathbf{d}(t)),\text{ a.e. } t\geq 0,\\ &\xi(0) = x,\end{aligned} \tag{2}$$

where $\xi(\cdot)$ describes the evolution of the dynamical system [6].

Obstacles in the environment can be modeled as a "keep-out" set of states $\mathcal{K}\subset\mathbb{R}^n$ that the system must avoid. We define the safety of the system with respect to this set, such that the system is considered to be safe at state $\xi(0)=x$ over time horizon T as long as we can choose $\mathbf{u}(\cdot)$ to guarantee that there exists no time $t\in[0,T]$ for which $\xi(t)\in\mathcal{K}$. The task of maintaining the system's safety over this interval can be modeled as a differential game between the control input and the disturbance. Consider an optimal control signal $\mathbf{u}(\cdot)$ which attempts to steer the system away from $\mathcal{K}$ and an optimal disturbance $\mathbf{d}(\cdot)$ which attempts to drive the system towards $\mathcal{K}$. By choosing any Lipschitz payoff function $l:\mathbb{R}^n\to\mathbb{R}$ which is negative-valued for $x\in\mathcal{K}$ and positive for $x\notin\mathcal{K}$, we can encode the outcome of this game via a value function $V(x,t)$ characterized by the following Hamilton-Jacobi-Isaacs variational inequality [8]:

$$\begin{aligned}&\min\begin{cases} l(x)-V(x,t),\\ \frac{\partial V}{\partial t}(x,t)+\max\limits_{u\in\mathcal{U}}\min\limits_{d\in\mathcal{D}}\frac{\partial V}{\partial x}(x,t)\cdot f(x,u,d)\end{cases} = 0\\ &V(x,T)=l(x).\end{aligned} \tag{3}$$

The value function $V(x,t)$ that satisfies the above conditions is equal to $\min_{\tau\in[t,T]} l(\xi^*(\tau))$ for the trajectory with $\xi^*(t)=x$ driven by an optimal control $\mathbf{u}(\cdot)$ and an optimal disturbance $\mathbf{d}(\cdot)$. We can therefore find the set of states $\mathcal{R}_T=\{x\in\mathbb{R}^n : V(x,0)<0\}$ from which we cannot guarantee the safety of the system on the interval $[0,T]$, also known as the backward-reachable set of $\mathcal{K}$ over this interval. That is, for all initial states $x\in\mathcal{R}_T$ and feedback control polices $\mathbf{u}(t)=g(\xi(t))$, there exists some disturbance $\mathbf{d}(\cdot)\in\mathscr{D}$ such that $\xi(t)\in\mathcal{K}$ for some $t\in[0,T]$.

If there exists a non-empty controlled-invariant set Ω that does not intersect $\mathcal{K}$, then we deem this set Ω a "safe set" because there exists a feedback policy that guarantees that the system remains in Ω, and thus out of $\mathcal{K}$, for all time. It follows from their properties that Ω is the complement of $\mathcal{R}_T$, and the relationship between $\mathcal{K}$, $\mathcal{R}_T$, and Ω is visualized in Fig. 2. Within a safe set Ω, the value function becomes independent of t as $T\to\infty$ [8]. Because we focus on the case where the system is initialized to some safe state $\xi(0)\in\Omega$ and we aim to maintain $\xi(t)\in\Omega$ for all $t\in[0,\infty)$, we simplify notation by defining the terms $V(x)\triangleq\lim_{T\to\infty}V(x,\cdot)$ and $\mathcal{R}\triangleq\mathcal{R}_\infty$.

One approach to guaranteeing the safety of the system is to apply a "minimally invasive" controller which activates on the zero level set of $V(x)$ [10]. This approach allows complete flexibility of control as long as $\xi(t)\in\text{interior}(\Omega)$, and applies the optimal control to avoid $\mathcal{K}$ when $\xi(\cdot)$ reaches the boundary of Ω. We refer the interested reader to [10], [8]

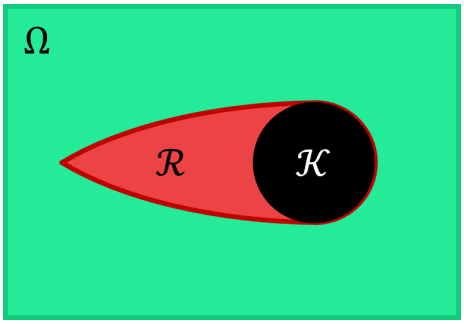

Fig. 2. Illustration of the relationship between a keep-out set $\mathcal{K}$, the derived backward-reachable set $\mathcal{R}$, and the resulting safe set Ω. Note that $\mathcal{K} \subseteq \mathcal{R}$, and Ω is equal to the complement of $\mathcal{R}$. This illustration approximates the result obtained using the Dubins car dynamics given in (10).

for a more thorough treatment of reachability and minimally invasive controllers.

B. Noisy Idealized Supervisor Model

We define an idealized model of the supervisor of a robotic team whose responsibility it is to ensure that no robots collide with obstacles represented by the keep-out set $\mathcal{K}$. The idealized supervisor behaves as a minimally invasive controller as described in Section II-A. However, while the robotic team members' true dynamics are given by the function $f(x, u, d)$ as in (1), the supervisor possesses an internal model of the robots' dynamics given by $f_S(x, u, d)$, which is not necessarily equal to the true dynamics. Following the differential game characterized by (3), the supervisor also possesses an internal value function $V_S(\cdot)$ and safe set Ω_S which they use to evaluate the safety of each state x in the environment. We allow for the possibility that the supervisor adds some amount of margin μ to their internal safe set, such that $\Omega_S = \{x \in \mathbb{R}^n : V_S(x) \geq \mu\}$. Therefore, the idealized supervisor will always intervene when a robotic team member reaches the μ level set of $V_S(\cdot)$, rather than the zero level set of the true $V(\cdot)$. We further specify that the idealized supervisor is *conservative*: $\forall x \in \mathbb{R}^n,\ V(x) \leq 0 \implies V_S(x) \leq \mu$. This condition implies that the supervisor will never let a robot teammate leave the true safe set Ω since $\Omega_S \subseteq \Omega$. We model our supervisor as a noisy version of this idealized supervisor. This noise is expressed as additive gaussian noise on $V_S(x)$ with zero mean and a standard deviation that is particular to the supervisor.

C. Learning Safe Sets from Supervisor Interventions

We choose to model the human supervisor of a robotic team as approximating the behavior of the idealized supervisor model presented in Section II-B. That is, the human supervisor will allow the robots to perform their task however they choose, but intervene whenever they *perceive* that a robot is approaching an obstacle $\mathcal{K}$ in the state space. Given this model, we can interpret the points at which the human intervenes as corresponding to the unknown μ level set of some value function $V_H(\cdot) : \mathbb{R}^n \to \mathbb{R}$, which characterizes the human's mental safe set Ω_H. Our goal is to use observations of human interventions to derive an estimated value function $\hat{V}_H(\cdot)$ and $\hat{\mu}$ which describe the observed behavior and induce an estimated $\hat{\Omega}_H$. We approach this task by deriving a Maximum Likelihood Estimator (MLE) of the human's mental safe set. If we assume that a human supervisor always intends to intervene at the μ level set of $V_H(x)$, but their ability to precisely intervene at this level is subject to Gaussian noise, either from observation error or variability in reaction time, then we can consider the value at an intervention point x_i as being drawn from a normal distribution centered at μ (that is, $V_H(x_i) \sim \mathcal{N}(\mu, \sigma^2)$).

Given a proposed value function $\hat{V}_H(\cdot)$ and a set of intervention points $[x_1, x_2, \cdots, x_p]$ with corresponding values $[\hat{V}_H(x_1), \hat{V}_H(x_2), \cdots, \hat{V}_H(x_p)]$, we wish to estimate the most likely μ and σ^2 to explain these interventions. Gaussian distributions induce the following probability density function for a single observation $\hat{V}_H(x_j)$

$$f\left(\hat{V}_H(x_j) \mid \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(\hat{V}_H(x_j) - \mu\right)^2}{2\sigma^2}\right) \tag{4}$$

which leads to the following probability density for a set of p independent observations

$$\begin{aligned} f\left(\hat{V}_H(x_1), \cdots, \hat{V}_H(x_p) \mid \mu, \sigma^2\right) &= \prod_{j=1}^{p} f\left(\hat{V}_H(x_j) \mid \mu, \sigma^2\right) \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{p}{2}} \exp\left(-\frac{\sum_{j=1}^{p}\left(\hat{V}_H(x_j) - \mu\right)^2}{2\sigma^2}\right). \end{aligned} \tag{5}$$

The likelihood of any estimated parameter values $\hat{\mu}$ and $\hat{\sigma}^2$ being correct, given the observations and the proposed value function $\hat{V}_H(\cdot)$, is expressed as $\mathcal{L}\left(\hat{\mu}, \hat{\sigma}^2 \mid \hat{V}_H(\cdot)\right) = f\left(\hat{V}_H(x_1), \cdots, \hat{V}_H(x_p) \mid \hat{\mu}, \hat{\sigma}^2\right)$. It can be shown that the values of the unknown parameters μ and σ^2 that maximize the likelihood function are given by

$$\hat{\mu}^* = \frac{1}{p}\sum_{j=1}^{p} \hat{V}_H(x_j) \quad \text{and} \quad \hat{\sigma}^{*2} = \frac{1}{p}\sum_{j=1}^{p}\left(\hat{V}_H(x_j) - \hat{\mu}^*\right)^2, \tag{6}$$

which are simply the mean and variance of the set of observations.

Notice that the estimates given by (6) are computed with respect to a given value function $\hat{V}_H(\cdot)$. If we were to assume that the human supervisor has a perfect model of the system dynamics, then we could simply set $\hat{V}_H(\cdot)$ to equal the true $V(\cdot)$ of the system in (1), and $\hat{\mu}^*$ would be the maximum likelihood estimate for the level at which the supervisor will intervene. However, it is unlikely that a human supervisor's notion of the dynamics will correspond exactly to this model, and we would like to maintain the flexibility of estimating value functions that are not strictly derived from (1). To this end, we define the maximum

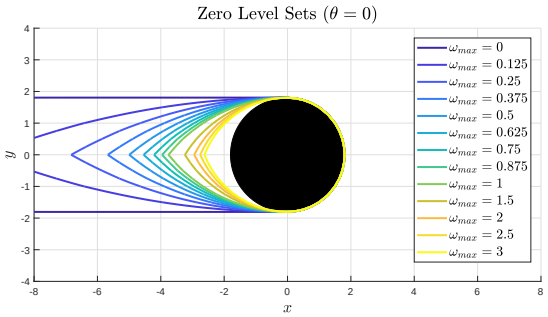

Fig. 3. Two dimensional slices of the zero level sets of the value functions $V_i(\cdot)$ from the library used for the experiment described in Section III. We used a family of Dubins car dynamics (see (10)) parametrized by ω_{max}. Notice that as ω_{max} decreases (the modeled control authority is decreased), the level sets extend farther away from the obstacle, indicating that a robot is expected to turn earlier to guarantee safety.

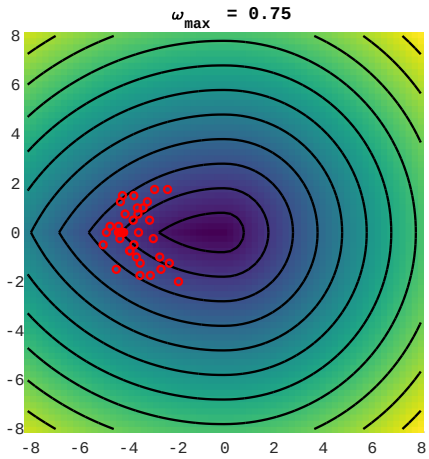

Fig. 4. An example data set from the experiment described in Section III. The red circles represent the location of supervisor interventions, and the colored background represents the learned value function $V(\cdot)$ with contour lines shown in black. In this case, the learning algorithm chose a dynamics model parametrized by $\omega_{max} = 0.75$.

likelihood of $\hat{V}_H(\cdot)$ being the $V_H(\cdot)$ that produced our observations as $\mathcal{L}^*(\hat{V}_H(\cdot)) = \max_{\hat{\mu},\hat{\sigma}^2} \mathcal{L}(\hat{\mu}, \hat{\sigma}^2 \mid \hat{V}_H(\cdot))$. The value of $\mathcal{L}^*(\hat{V}_H(\cdot))$ is obtained by substituting the estimates from (6) into the probability density function from (5). That is, $\mathcal{L}^*\left(\hat{V}_H(\cdot)\right) = f\left(\hat{V}_H(x_1), \cdots, \hat{V}_H(x_p) \mid \hat{\mu}^*, \hat{\sigma}^{*2}\right)$.

We seek the most likely value function to explain our observations, which will be the value function $\hat{V}^*(\cdot)$ with the greatest maximum likelihood $\mathcal{L}^*(\hat{V}^*(\cdot))$ (the maximum over maxima)

$$\hat{V}^*(\cdot) = \underset{V(\cdot)\in\mathcal{V}}{\arg\max}\, \mathcal{L}^*\left(V(\cdot)\right), \tag{7}$$

where $\mathcal{V}$ is the set of all possible value functions.

In order to make this optimization tractable, we can restrict ourselves to a set of value functions $\{V_\theta(\cdot)\}_{\theta\in\mathbb{R}^m}$ corresponding to a family of dynamics functions $\{f_\theta(\cdot,\cdot,\cdot)\}_{\theta\in\mathbb{R}^m}$ parameterized by $\theta \in \mathbb{R}^m$, making the optimization in question

$$\hat{V}^*(\cdot) = \underset{\theta\in\mathbb{R}^m}{\arg\max}\, \mathcal{L}^*\left(V_\theta(\cdot)\right). \tag{8}$$

In practice, we may not be able to find an expression for the gradient of $\mathcal{L}^*(V_\theta(\cdot))$ with respect to θ, since the value function is derived from the dynamics $f_\theta(\cdot,\cdot,\cdot)$ via the differential game given by (3). The lack of a gradient expression restricts the use of numerical methods to solve the problem as presented in (8). In these cases, we can compute a representative library of b value functions $\{V_i(\cdot)\}_{i=1}^b$ corresponding to a set of b representative parameter values $\{\theta_i\}_{i=1}^b$ (see Fig. 3 for an example library). The optimization then reduces to choosing the most likely value function from among this library

$$\hat{V}^*(\cdot) = \underset{i\in\{1,\cdots,b\}}{\arg\max}\, \mathcal{L}^*\left(V_i(\cdot)\right). \tag{9}$$

In order to ensure that the learned safe set is conservative, we can extend our MLE to a Maximum A Posteriori (MAP) estimator by incorporating our prior belief that, regardless of the safe set that the supervisor uses to generate interventions, they do not want the robots to be unsafe with respect to the true dynamics. In this case, we maintain a uniform prior $P(\theta)$ that assigns equal probability to all $V_\theta(\cdot)$ whose zero sublevel sets are supersets of the zero sublevel set of the true $V(\cdot)$, and zero probability to all other $V_\theta(\cdot)$. In other words, we assume that the supervisor does not overestimate the agility of the robots, and in practice we can enforce this condition by choosing the library in (9) to only contain appropriate value functions. Moreover, regardless of the choice of $\hat{V}_H(\cdot)$, we assume that the supervisor intends to intervene before reaching the zero level set of $\hat{V}_H(\cdot)$, which always includes the boundary of $\mathcal{K}$. If we choose a prior $P(\mu)$ that assigns zero probability to all non-positive μ and uniform probability elsewhere, it can be shown that the MAP estimates are obtained by letting $\hat{\mu}^*$ equal $\max\{\hat{\mu}^*, 0\}$ and otherwise proceeding as before. Fig. 4 provides an example of this algorithm estimating a safe set from human supervisor intervention data.

D. Team Control with Learned Safe Sets

We propose that safe sets learned according to the approach in Section II-C can be used to create effective control laws for the robotic members of human-robot teams. Recall our model of the human supervisor of a robotic team: the supervisor must rely on each robot's autonomy to complete the majority of their tasks unassisted, but the supervisor may intervene to correct a robot's behavior when necessary (such as by avoiding an imminent collision with the keep-out set $\mathcal{K}$). We put forth that in the scenario where the human intervenes to prevent a collision, they do so because they observe that a robot has violated the boundaries of their mental safe set Ω_H.

Now, consider a team of robots navigating an unknown

environment, and which are able to avoid any obstacles that they detect. One approach to safely automating this team is to have each robot behave according to a minimally invasive control law: the robots are allowed to follow trajectories generated by any planning algorithm, so long as they remain within Ω, the reachable set computed using the baseline dynamics model (1) with associated value function $V(\cdot)$. Whenever these robots detect an obstacle, they add it to the keep-out set $\mathcal{K}$, thus modifying Ω and $V(\cdot)$. If a robot reaches the boundary of Ω, it applies the optimal control to avoid $\mathcal{K}$ until it has cleared the obstacle. However, it is possible that a robot does not detect an obstacle, and a human supervisor must intervene to ensure robot safety.

As stated above, the human supervisor will intervene when a robot reaches the boundary of Ω_H, not the boundary of Ω. This discrepancy leads to the possibility that the supervisor will intervene when the robot reaches some state x, even if the robot would have avoided the obstacle without intervention. These situations arise whenever $V_H(x) \leq \mu$ but $V(x) > 0$. These "false positive" interventions represent unnecessary work for the human supervisor, and we seek to eliminate them in order to improve the human's experience and the team's overall performance.

We propose using a safe set $\hat{\Omega}_H$ learned from previous observations of supervisor interventions, as outlined in Section II-C, as a substitute for Ω in the robots' minimally invasive control law. By estimating the human's internal safe set, we take advantage of the following property:

Property. *For an idealized supervisor collaborating with a team of robots as described in Section II-D, if the robots avoid detected obstacles $\mathcal{K}$ by applying an optimally safe control at the boundary of safe set Ω_S, then if the supervisor plans to intervene because they observe $\xi_i(t) \in R_S$ for robot i, the supervisor can infer that robot i has not detected an obstacle and any supervisor intervention will not be a false positive.*

Proof. The proof of this property follows constructively from the definition of safe set, idealized supervisor, and false positive. If robot i had correctly detected an obstacle and adjusted its representation of Ω_S accordingly, then it would have applied the optimal control to remain within the supervisor's safe set. Therefore, if the supervisor is able to observe that robot i has left Ω_S, it must be the case that the robot has not detected the obstacle. False positives are defined to be supervisor interventions that occur when a robot had detected an obstacle but the supervisor still intervenes. In this case, the supervisor correctly infers that robot i has not detected an obstacle, so any intervention at this point cannot be a false positive. ■

For an idealized supervisor, as $\hat{\Omega}_H$ becomes an arbitrarily good approximation of Ω_H, the number of false positive interventions will approach zero. For a *noisy* idealized supervisor, the supervisor will intervene whenever $V_H(x) + w(x) \leq \mu$ where $w(x) \sim \mathcal{N}(0, \sigma_H^2)$. This noise will continue to produce false positives even with a perfect fit $\hat{\Omega}_H = \Omega_H$ if the robots apply the optimally safe control at the μ-level set of Ω_H. Instead, the level set α where the optimally safe control is applied can be raised arbitrarily high to drive the false positive rate to zero. For example, $\alpha = \mu + 2\sigma_H$ is sufficient to decrease the false positive to 5%. We test the efficacy of our approach through the human-subjects experiment described in Section III.

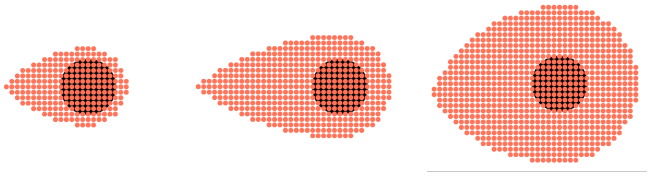

Fig. 5. Safe sets tested in user study: (left) physics-calculated safe set, (middle) example learned safe set, (right) extra conservative safe set

III. Experimental Design for User Validation

Our goal in understanding and modeling the supervisor's conception of safety is to improve team performance by decreasing cognitive overload. Although we have grounded our human modeling in the cognitive science literature, we do not claim to verify our humble extension. Instead, we aim to apply cognitive science toward building better human-robot teams. To this end, our hypotheses are:

1) H1: Representing supervisor behavior as cognitive keep-out sets allows intervention signals to be distilled into an actionable rule which will decrease supervisory false positives and cognitive strain, thereby increasing team performance and trust.

2) H2: Fitting danger-avoidance behavior to the supervisor's unique beliefs is preferable to generic conservative behavior.

In our experiment, we gather supervisor intervention data, fit our model to the data, and then runs a human-robot teaming task that assesses performance.

A. Procedure

Our experiment applies the idealized supervisor theory and learning algorithm to supervising self-driving cars. Experimental design benefits from simplification and eliminating the confounds inherent in complexity, therefore we did not work with physical autonomous cars, substituting an interactive simulation instead. The cars moved according to the Dubins car model:

$$\begin{aligned} \dot{x} &= 3\cos(\theta) \\ \dot{y} &= 3\sin(\theta) \\ \dot{\theta} &= u \\ u \in \mathcal{U} &= [-\omega_{max}, \omega_{max}],\ \omega_{max} = 1 \end{aligned} \tag{10}$$

The experiment is divided into three phases. First, the subject is given an opportunity to familiarize themselves with the robotic system's dynamics. The user is allowed to directly apply the full range of controls through the computer keyboard for one minute. After ensuring the user has some experience to build their cognitive dynamics from, we then

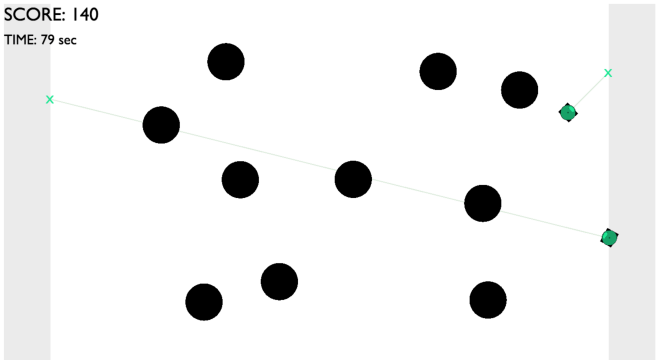

Fig. 6. Screenshot of experimental task for assessing supervisory success with learned safe set. Robotic vehicles make trips across the screen dodging obstacles with an 80% chance. The human supervisor operates as a failsafe for saving the robots in the 20% chance of not-detecting the obstacle.

assess the emergent conception of safety. In this second phase, supervisory data is extracted from the subject by showing them scenes where the car is driving towards an obstacle, and the supervisor decides where to intervene to avoid a crash. This intervention data is then fed into our algorithm (described in Sec. II-C) that extracts the best fitting safe set. In this experiment, we enforced conservativeness by excluding subjects whose learned sets were not supersets of the physics-based safe set, rather than enforce a prior on estimated turning rate. This safe set is assessed in the third phase alongside two pre-calculated baselines (see Fig. 5).

These safe sets were calculated using Hamilton-Jacobi reachability as described in Sec. II-A using the Level Set Toolbox [17] for MATLAB. During this final phase, the subject sequentially supervises homogeneous teams of robots, each team avoiding obstacles based on one of the three assessed safe sets. Ten randomly placed obstacles are strewn about the screen impeding the robots' autonomous trips back and forth across the screen (see Fig. 6). Although robots will detect and avoid an obstacles in 80% of their interactions with it, there is a 20% chance that the robot will not detect an obstacle as it approaches. The subject is charged with catching these random failures and removing an obstacle before the robot crashes. Crashing is disincentivized by decrementing an on-screen "score" counter. Removing an obstacle costs only half of what a crash costs the player. This encourages saving the robot but not guessing wildly and clearing out obstacles will not work as every obstacle removed spawns a new obstacle elsewhere. This "score" mechanism was also used to make the participant invested in team success, as every time a robot team-mate completed a trip across the screen their score increased.

B. Independent Variables

To assess our hypotheses, we manipulate the safe set used between team supervision trials. We exposed the human subject to three teams, each driving using one of three safe sets. The main kernel is learned from supervisor intervention signals as described in Section II-C. The two baseline kernels are calculated using Hamilton-Jacobi-Isaacs reachability on the true dynamic equations. One is calculated using the true obstacle size. The other adds a buffer that doubles the effective size of the obstacle, inducing extra-conservative behavior that gives obstacles a wide berth.

C. Dependent Measures

1) Objective Measures: The team was tasked with making trips across the screen to reach randomized goals. The robots' task was to travel across the screen, safely dodging obstacles along the way, while the human was tasked with supervising as a failsafe to remove an obstacle if the robots should fail to observe and avoid it.

Team performance was quantified using three objective metrics: number of trips completed, number of supervisory interventions, and the number of obstacle collisions. These metrics were presented to the subject as an arcade-style score that increased with each completed trip. To incentivize efficient supervision, obstacle-removal interventions cost the subject points, but only half as much as an obstacle collision would cost.

The number of interventions taken by the supervisor can also serve as a proxy measurement to quantify the amount of cognitive strain they experience while working with the robotic team. Of particular note are the number of interventions that were not actually required as the robot would have dodged the obstacle in time, but the supervisor was not convinced. These false positives needlessly drain supervisor attention and indicate a lack of trust in the system. We aim to increase the human's trust in the system, which we quantify by a decrease in these false positives.

2) Subjective Measures: After each round of pairwise comparison (completing the task with two different robotic teams), we presented the subject with a questionnaire to gauge how the choice of safe set impacted their experience. These questionnaires contained statements about each team that subjects would respond to using a 7-point Likert scale (1 - Strongly Disagree, 7 - Strongly Agree). These statements were designed to measure Trust, Perceived Performance, Interpretability, Confidence, Team Fluency, and overall Preference between the teams in the comparison.

D. Subject Allocation

The subject population consisted of 6 male, 5 female, and 1 non-binary participants between the ages of 18-29. We used a within-subjects design where each subject was asked to complete all three possible pairwise comparisons of our three treatments (the safe sets used). Within-subjects designs often suffer confounding effects from order. To ameliorate this we used a balanced Latin Square design for the order of comparisons with no treatment being first in a pair twice. Furthermore, we generated six randomized versions of the task so that subjects were presented with a different version of the task for each trial across the three pairwise comparisons. To avoid coupling the treatment results to a particular version of the task, each treatment was paired with each task version an equal number of times across our subject population.

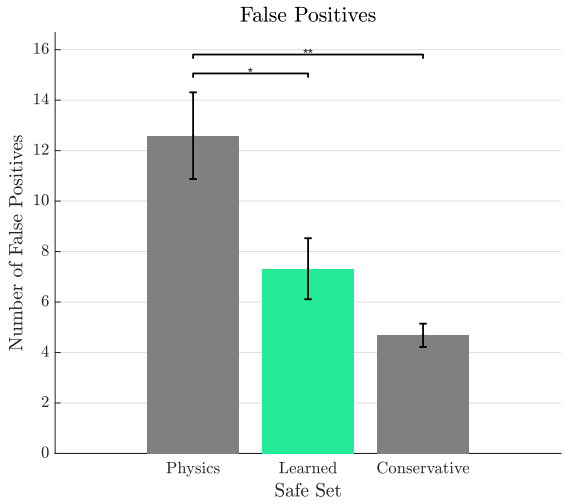

Fig. 7. False positives per trial plotted against the three safe set types. There were significant differences between physics and learned ($p < .05$) and between physics and conservative ($p < .01$). There was no significant difference between learned and conservative.

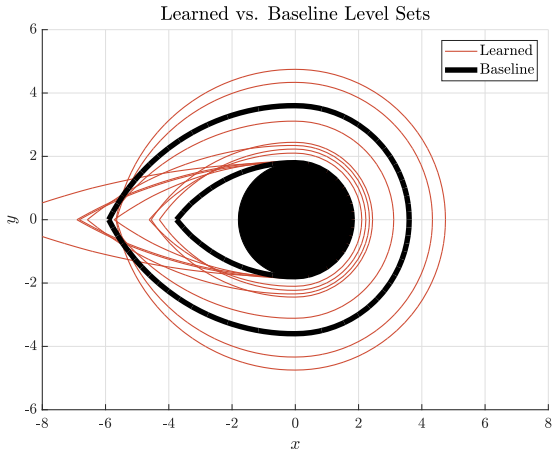

Fig. 8. Regressed safe sets (viewed on the $\theta = 0$ slice) from supervisor intervention data overlaid on physics-calculated baselines. Three users' safe sets clustered to arcing like the minimal physics safe set. Three others clustered to arcing like the conservative safe set. The final five safe sets exhibit a distinct behavior that reflects supervisors' preference for gradual, pre-emptive arcs.

IV. Analysis and Discussion

A. H1: False Positives versus Standard Interventions

Every supervisor intervention costs cognitive resources. Unnecessary interventions (where the supervisor perceives a danger where the system is actually safe) represent wasted team potential. Our first hypothesis is that an expanded safe set that reflects the supervisor's intervention behavior would decrease the number of false positives. To test this, we performed a one-way repeated measures ANOVA on the number of supervisory false positives from the third phase of the experiment with safe set as the manipulated factor. A "false positive" was any supervisor intervention where the obstacle was actually detected by all nearby robots which would have avoided the obstacle successfully. The robot team's safe set had a significant effect on the number of false supervisory positives ($F(2, 20) = 8.72, p < 0.01$). An all-pairs post-hoc Tukey method found that the learned safe set significantly decreased ($p = 0.0328 < 0.05$) false positives over the standard physics-calculated safe set, but there was no significant difference between the learned safe set and the conservative safe set (which also significantly decreased false positives over the physics-calculated safe set with $p < 0.01$). These results support our main hypothesis that ***representing supervisor behavior as cognitive keep-out sets allows intervention signals to be distilled into an actionable rule which will decrease supervisory false positives***.

The second half of that hypothesis, that ***decreasing supervisory false positives will increase trust and team performance*** was not shown conclusively from our data. We performed a post-hoc, one-way, repeated measures ANOVA on the pairwise comparison surveys between the teams utilizing the learned and the physics-calculated safe sets. Measures of Trust showed no significant improvement ($F(1, 9) = 1.86, p = 0.21$).

B. H2: Preference versus Conservative Interventions

For 9 of 11 participants the learned safe set had shorter avoidance arcs than the conservative set. We hypothesized that this greater efficiency would make the tailored conservativeness of the learned set preferable to the generic safe set, yet results showed no significant difference in the preference score between the two. A t-test showed that the preference was statistically indistinguishable ($p = 0.8$) from a neutral score: an inconclusive result for Hypothesis 2. We believe that this indistinguishability stems from users judging preference more on intelligibility than on efficiency. The supervisor model predicts that false positives will significantly decrease for larger safe sets like the conservative set used in this experiment (see next section on Model Validity). This significant decrease was also observed experimentally as mentioned above.

This indistinguishability is further compounded since a preference for intelligibility seems to be expressed in the supervisory intervention data, resulting in learned safe sets with similar arcs as the conservative safe set (see Fig. 8). Future work could investigate this efficiency-intelligibility tradeoff further by using a conservative baseline that is distinguishably more conservative than user safe sets and making efficiency more central to the team task.

C. Model Validity

The statistically significant decreases in false positives observed in Sec. IV-A match the decreases predicted by the supervisor model. Our supervisor model posits that intervention behavior is tied to states noisily distributed about a cognitive safe set. It would predict that the empirical distribution of intervention states inside a proposed safe set (as illustrated in Fig. 9) will mirror the number of

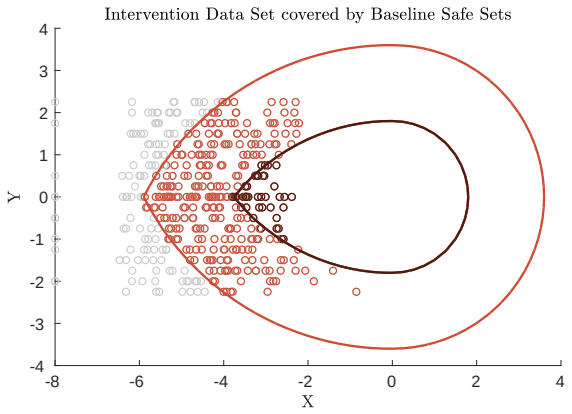

Fig. 9. Empirical distribution of flinch states observed during data collection (Phase 2 of experiment). The 325 interventions covered by the conservative baseline safe set are colored in red, with the 43 interventions covered by the physics-calculated safe set are colored darker. All intervention states covered by a baseline set will be avoided when that set is used in the human-robot teaming task.

	Interventions	Dec (%)	Avg. F.P.	Dec (%)
Physics-based	397 / 440	—	12.54	—
Learned	220 / 440	44.4	7.31	41.7
Conservative	115 / 440	71	4.68	62.7

Table 1: Model predictions for decrease in interventions compared to decrease in false positives. Left: Predicted reduction in interventions from training data over choice of safe sets. Right: Measured decrease in false positive interventions from human-robot teaming task.

interventions prevented. Since the learned safe set level α is chosen as the mean $\hat{\mu}$ of the Gaussian generative model for supervisory behavior (see Sec. II-C), exactly half the distribution will be covered by the learned set. This model prediction is compared in Tbl. 1 to the observed decrease in false positive interventions.

V. Conclusion

Automation with human supervisors relies on leveraging the human supervisor's cognitive resources for success. Respecting these resources is essential for creating well-performing human-robot teams. It is especially important to avoid overtaxing the human as automated-teams continue to scale up, and a single human worker both accomplishes more and bears more cognitive load than ever. To alleviate cognitive load, we can decrease the quantity of issues that command the supervisor's attention by pruning false positives. By modeling what system states command supervisory attention, we can program robots to avoid those areas when they don't need attention. We constructed such a model for human supervisory behavior by combining notions of safety from reachability theory with mental-simulation models in cognitive science. We demonstrated the model's efficacy on understanding and decreasing supervisory false positives in a human-robot teaming task.

References

[1] Pieter Abbeel and Andrew Y Ng. Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the twenty-first international conference on Machine learning*, page 1. ACM, 2004.

[2] Anayo K Akametalu and Claire J Tomlin. Temporal-difference learning for online reachability analysis. In *Control Conference (ECC), 2015 European*, pages 2508–2513. IEEE, 2015.

[3] Andrea Bajcsy and Anca Dragan. Learning robot objectives from physical human interaction, stand-in entry. In *International Conference on Robotics and Automation (ICRA) 2018*, pages 9001–10200. IEEE, 2018.

[4] Peter W Battaglia, Jessica B Hamrick, and Joshua B Tenenbaum. Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, 110(45):18327–18332, 2013.

[5] Mo Chen, Jaime F Fisac, Shankar Sastry, and Claire J Tomlin. Safe sequential path planning of multi-vehicle systems via double-obstacle hamilton-jacobi-isaacs variational inequality. In *Control Conference (ECC), 2015 European*, pages 3304–3309. IEEE, 2015.

[6] Earl A Coddington and Norman Levinson. *Theory of ordinary differential equations*. Tata McGraw-Hill Education, 1955.

[7] Anca D Dragan and Siddhartha S Srinivasa. A policy-blending formalism for shared control. *The International Journal of Robotics Research*, 32(7):790–805, 2013.

[8] Jaime F. Fisac, Anayo K. Akametalu, Melanie Nicole Zeilinger, Shahab Kaynama, Jeremy H. Gillula, and Claire J. Tomlin. A general safety framework for learning-based control in uncertain robotic systems. *CoRR*, abs/1705.01292, 2017.

[9] Jaime F Fisac, Mo Chen, Claire J Tomlin, and S Shankar Sastry. Reach-avoid problems with time-varying dynamics, targets and constraints. In *Proceedings of the 18th international conference on hybrid systems: computation and control*, pages 11–20. ACM, 2015.

[10] Jeremy H. Gillula, Gabriel M. Hoffmann, Haomiao Huang, Michael P. Vitus, and Claire J. Tomlin. Applications of hybrid reachability analysis to robotic aerial vehicles. *The International Journal of Robotics Research*, 30(3):335–354, 2011.

[11] Sylvia L Herbert, Mo Chen, SooJean Han, Somil Bansal, Jaime F Fisac, and Claire J Tomlin. Fastrack: a modular framework for fast and guaranteed safe motion planning. *arXiv preprint arXiv:1703.07373*, 2017.

[12] Gabriel M Hoffmann and Claire J Tomlin. Decentralized cooperative collision avoidance for acceleration constrained vehicles. In *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, pages 4357–4363. IEEE, 2008.

[13] Sheue-Ling Hwang, Woodrow Barfield, Tien-Chen Chang, and Gavriel Salvendy. Integration of humans and computers in the operation and control of flexible manufacturing systems. *The International Journal of Production Research*, 22(5):841–856, 1984.

[14] Ashesh Jain, Shikhar Sharma, Thorsten Joachims, and Ashutosh Saxena. Learning preferences for manipulation tasks from online coactive feedback. *The International Journal of Robotics Research*, 34(10):1296–1313, 2015.

[15] Shervin Javdani, Henny Admoni, Stefania Pellegrinelli, Siddhartha S Srinivasa, and J Andrew Bagnell. Shared autonomy via hindsight optimization for teleoperation and teaming. *arXiv preprint arXiv:1706.00155*, 2017.

[16] Rudolf Emil Kalman. When is a linear control system optimal? *Journal of Basic Engineering*, 86(1):51–60, 1964.

[17] Ian M Mitchell. A toolbox of level set methods. *Dept. Comput. Sci., Univ. British Columbia, Vancouver, BC, Canada, http://www. cs. ubc. ca/˜ mitchell/ToolboxLS/toolboxLS. pdf, Tech. Rep. TR-2004-09*, 2004.

[18] Kevin A Smith and Edward Vul. Sources of uncertainty in intuitive physics. *Topics in cognitive science*, 5(1):185–199, 2013.

[19] Claire J Tomlin. Towards automated conflict resolution in air traffic control1. *IFAC Proceedings Volumes*, 32(2):6564–6569, 1999.

[20] Brian D Ziebart, Andrew L Maas, J Andrew Bagnell, and Anind K Dey. Maximum entropy inverse reinforcement learning. In *AAAI*, volume 8, pages 1433–1438. Chicago, IL, USA, 2008.