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 collaborates with a team of robots, the human’s attention is divided, and cognitive resources are at a premium. We aim to optimize the distribution of these resources and the flow of attention. To this end, we propose the model of an idealized supervisor to describe human behavior. Such a supervisor employs a potentially inaccurate internal model of the robots’ dynamics to judge safety. We represent these safety judgements by constructing a safe set from this internal model using reachability theory. When a robot leaves this safe set, the idealized supervisor will intervene to assist, regardless of whether or not the robot remains objectively safe. False positives, where a human supervisor incorrectly judges a robot to be in danger, needlessly consume supervisor attention. In this work, we propose a method that decreases false positives by learning the supervisor’s safe set and using that information to govern robot behavior. We prove that robots behaving according to our approach will reduce the occurrence of false positives for our idealized supervisor model. Furthermore, we empirically validate our approach with a user study that demonstrates a significant ($p = 0.0328$) reduction in false positives for our method compared to a baseline safety controller.

I. INTRODUCTION AND BACKGROUND

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 human workers, since it will become possible for a single human to supervise the behavior of multiple robotic agents.

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 [9], [16]). 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.

This project is inspired by work like Bajcsy et al [3] and Jain et al [15] 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. The authors of [3] posed this correction challenge in model-based framework that interprets the human’s signals as resulting from an optimization problem. This inverse optimization framework has also been used in Inverse Reinforcement Learning [1], [20] which applies Inverse Optimal Control (as conceived of by Kalman [17]) to interpreting human trajectories. Our work applies the 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], [19]. We posit that human supervisors utilize this same cognitive dynamic simulation to predict robot safety and intervene accordingly. Specifically, we theorize that the intervention behavior is driven by an internal “safe set” which we can attempt to reconstruct by observing supervisor interventions.

Safe sets are conceived from 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 first two authors contributed equally

This work is supported by the Office of Naval Research under the Embedded Humans MURI (N00014-13-1-0341) as well as a Philippine-California Advanced Research Institutes (PCARI) grant.

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

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.
the set and therefore stay safely away from the undesirable states. For this reason, viability kernels are often referred to as “safe sets”. Reachability can be used for robust path planning in dynamically changing environments [11] 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 [13].

Hoffman et al. used the safety guarantees of reachability analysis to engineer a multi-drone team that could automatically avoid collisions [14]. Similarly, Gillula could guarantee safety for learning algorithms by constraining their explorations to stay within the safe set [12]. 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

$$\dot{x} = f(x, u, d),$$

$x \in \mathbb{R}^n$, $u \in \mathcal{U} \subset \mathbb{R}^n$, $d \in \mathcal{D} \subset \mathbb{R}^n$, \hspace{1cm} (1)$

where $\mathcal{U}$ and $\mathcal{D}$ are compact. We let $\mathcal{U}$ and $\mathcal{D}$ denote the sets of measurable functions $u : [0, \infty) \rightarrow \mathcal{U}$ and $\mathbf{d} : [0, \infty) \rightarrow \mathcal{D}$, 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 $\xi : [0, \infty) \rightarrow \mathbb{R}^n$ from any initial state $x$ which solves

$$\dot{\xi}(t) = f(\xi(t), u(t), \mathbf{d}(t)), \quad \text{a.e. } t \geq 0,$$

$$\xi(0) = x,$$ \hspace{1cm} (2)

where $\xi(\cdot)$ describes the evolution of the dynamical system [8].

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 $u(\cdot)$ to guarantee that there exists no time $t \in [0, T]$ for which $\xi(t) \notin \mathcal{K}$. 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 $u(\cdot)$ 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 $l : \mathbb{R}^n \rightarrow \mathbb{R}$ 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 [10]:

$$\min \left\{ l(x) - V(x, t), \right.$$

$$\left. \frac{\partial V}{\partial x}(x, t) + \max_{u \in \mathcal{U}} \min_{d \in \mathcal{D}} \frac{\partial V}{\partial x}(x, t) \cdot f(x, u, d) \right\} = 0,$$

$$V(x, T) = l(x).$$ \hspace{1cm} (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 $u(\cdot)$ 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 $u(t) = g(\xi(t))$, there exists some disturbance $\mathbf{d}(\cdot) \in \mathcal{D}$ such that $\xi(t) \notin \mathcal{K}$ for some $t \in [0, T]$.

If there exists a non-empty controlled-invariant set $\Omega$ that does not intersect $\mathcal{K}$, then we deem this set $\Omega$ a “safe set” because there exists a feedback policy that guarantees that the system remains in $\Omega$, and thus out of $\mathcal{K}$, for all time. It follows from their properties that $\Omega$ is the complement of $\mathcal{R}_T$, and the relationship between $\mathcal{K}$, $\mathcal{R}_T$, and $\Omega$ is visualized in Fig. 2. Within a safe set $\Omega$, the value function becomes independent of $t$ as $T \rightarrow \infty$ [10]. 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 $V(x) \triangleq \lim_{T \rightarrow \infty} V(x, \cdot)$ 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 $V(x)$ \[12\]. 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 $\Omega$. We refer the interested reader to \[12\], \[10\] 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 $f_\omega(x, u, d)$, which is not necessarily equal to the true dynamics. Following the differential game characterized by (3), the supervisor also possesses an internal value function $V_\omega(\cdot)$ and safe set $\Omega_\omega$ 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 $\mu$ to their internal safe set, such that $\Omega_\omega = \{x \in \mathbb{R}^n : V_\omega(x) \geq \mu\}$. Therefore, the idealized supervisor will always intervene when a robotic team member reaches the $\mu$ level set of $V_\omega(\cdot)$, rather than the zero level set of the true $V(\cdot)$. We further specify that the idealized supervisor is conservative: $\forall x \in \mathbb{R}^n$, $V(x) \leq 0 \implies V_\omega(x) \leq \mu$. This condition implies that the supervisor will never let a robot teammate leave the true safe set $\Omega$ since $\Omega_\omega \subseteq \Omega$. Additionally, we propose a noisy version of this idealized supervisor: the noisy idealized supervisor will intervene when they observe a robot reach the $\mu + w$ level set of $V_\omega(\cdot)$, where $w$ is drawn from $\mathcal{N}(0, \sigma^2_w)$ whenever a supervisor makes a safety judgement.

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 $\mu$ level set of some value function $V_H(\cdot) : \mathbb{R}^n \to \mathbb{R}$, which characterizes the human’s mental safe set $\Omega_H$. 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 $\mu$ 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 $\mu$ (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 interventions $\{x_1, x_2, \cdots, x_p\}$ 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 $\mu$ and $\sigma^2$ to explain these interventions. Gaussian distributions induce the following probability density function for a single observation $\hat{V}_H(x_j)$

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

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

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

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

$$L \left( \hat{\mu}, \hat{\sigma}^2 \mid \hat{V}_H(\cdot) \right) = f \left( \hat{V}_H(x_1), \cdots, \hat{V}_H(x_p) \mid \hat{\mu}, \hat{\sigma}^2 \right).$$

It can be shown that the values of the unknown parameters $\mu$ and $\sigma^2$ that maximize the likelihood function are given by

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

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 likelihood of $\hat{V}_H(\cdot)$ being the $V_H(\cdot)$ that produced our observations as

$$L^* \left( \hat{V}_H(\cdot) \mid \hat{\mu}^*, \hat{\sigma}^2 \right) \geq \max_{\mu, \sigma^2} L \left( \hat{\mu}, \hat{\sigma}^2 \mid \hat{V}_H(\cdot) \right).$$

The value of $L^* \left( \hat{V}_H(\cdot) \right)$ is obtained by substituting the estimates from (6) into the probability density function from (5). That is,

$$L^* \left( \hat{V}_H(\cdot) \right) = f \left( \hat{V}_H(x_1), \cdots, \hat{V}_H(x_p) \mid \hat{\mu}^*, \hat{\sigma}^2 \right).$$

We seek the most likely value function to explain our observations, which will be the value function $\tilde{V}_* (\cdot)$ with the greatest maximum likelihood $L^* \left( \tilde{V}_* (\cdot) \right)$ (the maximum
Whenever these robots detect an obstacle, they add it to the safe set that the supervisor uses to generate interventions, enforcing this condition by choosing the library in (9) to only contain appropriate value functions. Moreover, regardless of the choice of $V_H(\cdot)$, we assume that the supervisor intends to intervene before reaching the zero level set of $V_H(\cdot)$, which always includes the boundary of $K$. If we choose a prior $P(\mu)$ that assigns zero probability to all non-positive $\mu$ and uniform probability elsewhere, it can be shown that the MAP estimates are obtained by letting $\hat{\mu}^* = \max \{\hat{\mu}^*, 0\}$ 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 $K$). 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 $\Omega_H$.

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 $\Omega$, 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

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 $V_H(\cdot)$, we assume that the supervisor intends to intervene before reaching the zero level set of $V_H(\cdot)$, which always includes the boundary of $K$. If we choose a prior $P(\mu)$ that assigns zero probability to all non-positive $\mu$ and uniform probability elsewhere, it can be shown that the MAP estimates are obtained by letting $\hat{\mu}^* = \max \{\hat{\mu}^*, 0\}$ 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 $K$). 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 $\Omega_H$.

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 $\Omega$, 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 $K$, thus modifying $\Omega$ and $V(\cdot)$. If a robot reaches
the boundary of $\Omega$, it applies the optimal control to avoid $K$
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 $\Omega_H$, not the boundary
of $\Omega$. 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 $\Omega$ 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 $K$ by applying an optimally safe
control at the boundary of safe set $\Omega_S$, 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 definitions of safe set, idealized supervisor, and
false positive. If robot $i$ had correctly detected an obstacle
and adjusted its representation of $\Omega_S$, 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 $\Omega_S$, 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
has detected an obstacle but the supervisor still intervenes.
In this case, the supervisor can correctly infer that robot $i$
has not detected an obstacle, so any intervention at this point
cannot be a false positive.

For an idealized supervisor, as $\hat{\Omega}_S$ becomes an arbitrarily
good approximation of $\Omega_S$, the number of false positive inter-
ventions will approach zero. For a noisy idealized supervisor,
the supervisor will intervene whenever $V_S(x) + w \leq \mu$
where $w \sim \mathcal{N}(0, \sigma_S^2)$. This noise will continue to produce false
positives, even with a perfect fit $\hat{\Omega}_S = \Omega_S$, if the robots
apply the optimally safe control at the $\mu$-level set of $\Omega_S$.
Instead, the level set $\alpha$ where the optimally safe control
is applied can be raised arbitrarily high to drive the false
positive rate to zero. For example, $\alpha = \mu + 2\sigma_S$ is sufficient
to avoid over 97% of intervention states used for learning,
in expectation. We test the efficacy of our approach through
the human-subjects experiment described in Section III.

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 based our
human modeling on the cognitive science literature, we
do not intend to verify humans’ exact cognitive processes.
Instead, we aim to apply our inspiration from 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 per-
formance and trust.

2) $H2$: Fitting danger-avoidance behavior to a supervi-
sor’s beliefs is preferable to generic conservative behavior.

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

A. Procedure

Our experiment applies the idealized supervisor theory
and learning algorithm to supervising simulated robots. The
robots moved according to the Dubins car model:

$$\begin{align*}
\dot{x} &= 3 \cos(\theta) \\
\dot{y} &= 3 \sin(\theta) \\
\dot{\theta} &= u
\end{align*}$$

(10)

$u \in \mathcal{U} = [-\omega_{\max}, \omega_{\max}], \omega_{\max} = 1$

The experiment is divided into three phases. In Phase I,
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 from which to build an internal dynamics model,
we then assess their emergent conception of safety. In
Phase II, supervisory data is extracted from the subject by
showing them scenes where the robot 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 Section II-C) that extracts the best
fitting safe set. Our estimator used a library of candidate
dynamics functions parameterized by values of $\omega_{\max}$ be-
tween 0 and 3, as shown in Fig. 3. In this experiment,

Fig. 5. Safe sets tested in our experiment (illustrated by their complement-
tary reachable set): (left) Standard safe set (calculated from true dynamics
and obstacle size), (middle) example Learned safe set (calculated from fitted
supervisory perception of dynamics and obstacle size), (right) Conservative
safe set (calculated from true dynamics and inflated obstacle size)

865
we enforced conservativeness by excluding subjects whose Learned sets were not supersets of the Standard safe set, rather than enforcing a prior directly on $\omega_{\text{max}}$. The Learned safe set is assessed in Phase III against two fixed safe sets (see Fig. 5) pre-calculated from the true dynamic equations.

These safe sets were calculated using Hamilton-Jacobi reachability as described in Section II-A using the Level Set Toolbox [18] 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 disincetivized by decrementing an on-screen “score” counter. Removing an obstacle costs only half of what a crash costs the player. This system encourages saving the robot but not guessing wildly. Moreover, simply clearing out all obstacles is not a viable strategy because every obstacle removed generates a new obstacle elsewhere. This score mechanism was also used to make the participant invested in team success by awarding points every time a robot completes a trip across the screen.

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 Learned set is derived from Phase II supervisor intervention observations as described in Section II, using $\alpha = \mu$. The two baseline kernels are calculated using Hamilton-Jacobi-Isaacs reachability on the true dynamic equations. The Standard set is calculated using the true obstacle size. The Conservative set adds a buffer that doubles the effective size of the obstacle, inducing trajectories that give 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 aggregated, arcade-style score. To incentivize participants to only intervene when necessary, obstacle-removal interventions reduced the score, but only by half as much an obstacle collision.

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 supervisor incorrectly judged that a robot had not detected an obstacle. 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). 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.

IV. ANALYSIS AND DISCUSSION

A. H1: False Positive Reduction over Standard

Our first hypothesis is that a Learned safe set that reflects the supervisor’s intervention behavior would decrease the number of false positives compared to the Standard safe set. To test this, we performed a one-way repeated measures ANOVA on the number of supervisory false positives from
Phase III of the experiment with safe set as the manipulated factor. A false positive was any supervisor intervention where the removed obstacle was actually detected by all nearby robots, which would have avoided it successfully. The robot team’s safe set had a significant effect on the number of supervisory false positives \(F(2,20) = 8.72, p < 0.01\). An all-pairs post-hoc Tukey method found that the Learned safe set significantly decreased false positives over the Standard 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 Standard 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 one-way, repeated measures ANOVA on the pairwise comparison surveys between the teams using the Learned and the Standard safe sets. Measures of trust showed no significant improvement \(F(1,9) = 1.86, p = 0.21\).

B. H2: Preference over Conservative

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 baseline Conservative safe set. However, a t-test showed that the survey responses for preference were statistically indistinguishable \((p = 0.8)\) from a neutral score: an inconclusive result for Hypothesis 2. We believe that this result stems from users judging preference more on intelligibility, the ease of avoiding false positives, than on efficiency, the shortness of paths. As discussed in Section IV-A, both the Learned and Conservative safe sets led to significant false positive reductions over the Standard set.

This indistinguishability is further compounded since a preference for intelligibility seems to be expressed by some subjects in their Phase II intervention data, resulting in their Learned safe sets having similar arcs as the Conservative safe set (see Fig. 8). Future work could investigate this efficiency-intelligibility trade-off further by using a conservative baseline that is distinguishably more conservative than user safe sets and by making efficiency more central to the team task.

C. Model Validity

The statistically significant decreases in false positives observed in Phase III agree with the decreases predicted by the supervisor model based on intervention data from Phase II. Our model posits that interventions occur at states noisily distributed about a safe set boundary. Therefore, it predicts that the empirical distribution of Phase II intervention states contained within a proposed safe set (see Fig. 9) will mirror the proportion of false positive interventions observed in Phase III: if states are deemed safe by the controller, they will not be avoided, even when the noisy supervisor would judge them to be unsafe. Since the Learned safe set controller intervenes at the \(\hat{\mu}^+\) level set (see Section II-C), exactly half the intervention states will be contained within the Learned safe set in expectation. The model’s predictions are compared against observed false positives in Table 1.

| Safe Set | Interventions in Safe Set | Predicted vs Std. | Average F.P. | Observed F.P. |
|----------|---------------------------|-------------------|--------------|--------------|
| Standard | 397 / 440                 | 100%              | 12.54        | 100%         |
| Learned  | 220 / 440                 | 55.4%             | 7.31         | 58.3%        |
| Conservative | 115 / 440       | 29%               | 4.68         | 37.3%        |

Table 1: Predicted and observed false positives. Left: Predicted false positives from Phase II data. Right: Observed false positives in Phase III.
Our results show that it is possible to reduce false positives, and thus cognitive load, by aligning robot behavior with humans’ expectations. Future work will explore the impact of this framework on application domains from air traffic management to self-driving vehicles.

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 European Control Conference (ECC), pages 2508–2513. IEEE, 2015.
[3] Andrea Bajcsy, Dylan P Losey, Marcia K OMalley, and Anca D Dragan. Learning robot objectives from physical human interaction. In Conference on Robot Learning, pages 217–226, 2017.
[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-isacs variational inequality. In European Control Conference (ECC), pages 3304–3309. IEEE, 2015.
[6] Mo Chen, Sylvia Herbert, and Claire J Tomlin. Fast reachable set approximations via state decoupling disturbances. In Conference on Decision and Control (CDC), 2016, pages 191–196. IEEE, 2016.
[7] Mo Chen, Sylvia Herbert, and Claire J Tomlin. Exact and efficient hamilton-jacobi guaranteed safety analysis via system decomposition. In Robotics and Automation (ICRA), 2017 IEEE International Conference on, pages 87–92. IEEE, 2017.
[8] Earl A Coddington and Norman Levinson. Theory of ordinary differential equations. Tata McGraw-Hill Education, 1955.
[9] Anca D Dragan and Siddhartha Srinivasa. A policy-blending formalism for shared control. The International Journal of Robotics Research, 32(7):790–805, 2013.
[10] 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.
[11] Jaime F Fisac, Mo Chen, Claire J Tomlin, and 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.
[12] 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.
[13] S. L. Herbert, M. Chen, S. Han, S. Bansal, J. F. Fisac, and C. J. Tomlin. Fastrack: A modular framework for fast and guaranteed safe motion planning. In 2017 IEEE 56th Annual Conference on Decision and Control (CDC), pages 1517–1522, Dec 2017.
[14] Gabriel M Hoffmann and Claire J Tomlin. Decentralized cooperative collision avoidance for acceleration-constrained vehicles. In Conference on Decision and Control, pages 4357–4363. IEEE, 2008.
[15] Ashesh Jain, Shikhar Sharma, Thorsten Joachims, and Ashutosh Saxena. Learning preferences for manipulation tasks from online coactive feedback. The International Journal of Robotics Research, 32(10):1296–1313, 2013.
[16] Shervin Javdani, Henny Admoni, Stefania Pellegrinelli, Siddhartha S. Srinivasa, and J. Andrew Bagnell. Shared autonomy via hindsight optimization for teleoperation and teaming. The International Journal of Robotics Research, 37(7):717–742, 2018.
[17] Rudolf Emil Kalman. When is a linear control system optimal? Journal of Basic Engineering, 86(1):51–60, 1964.
[18] 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.
[19] Kevin A Smith and Edward Vul. Sources of uncertainty in intuitive physics. Topics in cognitive science, 5(1):185–199, 2013.
[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.