IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

Learning User Preferences for Complex Cobotic Tasks: Meta-Behaviors and Human Groups

Elena M. Vella[1], Airlie Chapman[1], Nir Lipovetzky [2]

***Abstract*— In complex tasks (beyond a single targeted controller) requiring robots to collaborate with multiple human users, two challenges arise: complex tasks are often composed of multiple behaviors which can only be evaluated as a collective (a meta-behavior) and user preferences often differ between individuals, yet successful interactions are expected across groups. To address these challenges, we formulate a set-wise preference learning problem, and validate a cost function that captures human group preferences for complex collaborative robotic tasks (cobotics). We develop a sparse optimization formulation to introduce a *distinctiveness* metric that aggregates individuals with similar preference profiles. Analysis of anonymized unlabelled preferences provides further insight into group preferences. Identification of the mode average most-preferred meta-behavior and minimum covariance bound allows us to analyze *group cohesion*. A user study with 43 participants is used to validate group preference profiles.**

***Index Terms*— Human-Robot Interaction, Control Design, Machine Learning Algorithms**

I. Introduction

The maturity of robotic systems has advanced to a point where humans and robots are able to work in close proximity. This has led to an increased prevalence of cobotic (collaborative robot) applications in industrial and social settings. Cobotic applications are characterized by interactions where humans and robots work together in teams to successfully complete the relevant task [1]. In this context, robots must be able to perform complex tasks alongside humans, where it is required of the robots to be able to execute coordinated motions beyond a single targeted controller, or behavior. To ensure successful group collaboration for complex tasks, a meta-behavior consisting of a set of dynamic motions for a complex task is required [2]. The intricate challenges of group collaboration in dynamic tasks involve negotiation, coordination, conflict, and social behavior. The inherent subjectivity of user criteria for cobotics, leads to user preferences being unavoidable in control parameterization design. It is pivotal to design cost functions that embody the subjective criteria of group preferences, that enable multiple human users and robots to collaborate.

To address the challenge of modeling user preference bias, it is necessary to establish an empirical measurement to design a cost function that is representative of human preferences. The idea of learning cost or reward functions from user-expressed preferences have been studied in the machine learning community at large [3]. The two primary approaches to preference learning are instance preference learning ordinal regression [4], [5] and *collections of comparisons* commonly seen as pairwise comparisons [6], [7]. In our work, we extend this to set-wise comparisons or comparing meta-behavior preferences. Mindful of the inherent subjectivity in user opinions, we capture individuals' preferences, such that the user provides a ranking, as opposed to absolute values. Inspired by the algorithm presented in [8], we use preferential reasoning to understand the influence of complex tasks achieved by a meta-behavior on human user's preferences.

In complex situations where multiple users engage with the robotic system, often the cost function represents a single user preference [9], resulting in poor performance over the group. We propose a cost function constructed to accurately represent a group's favored preferences, also referred to as a group preference profile. This is done by extending the concept of individual preferences to group preferences [10], thus providing a way of understanding how to aggregate individuals with similar preference profiles. We extend our analysis to consider the impact of conflicting preferences' in a group and how this affects the overall groups preference dynamic.

In this work, we seek to address, in complex cobotic tasks, how to design and interpret a group of human preference profiles to meta-behaviors. Our contributions extend concepts developed in [8], [10]–[12] to group based measurements for preferences. We apply a set-wise preference learning technique within a convex optimization framework. We consider a human group's preferences from two perspectives. First, we describe a population's *distinctiveness* by aggregating individuals with similar preferences. Secondly, we examine *group cohesion* via the notion of a preference average, and covariance bounds for the group's preference profiles. Finally, a user study of 43 individuals is conducted to verify user group preference profiles to meta-behaviors.

The remainder of this paper elaborates on our analytical approach to modeling human preferences and interpreting group preferences regarding meta-behaviors. We formulate

Manuscript received November 21, 2022; Revised February 13, 2022; Accepted May 7, 2023

This paper was recommended for publication by Venture, Gentiane upon evaluation of the Associate Editor and Reviewers' comments.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by (Office of Research Ethics and Integrity at the University of Melbourne) (ref2023-22472-20996-4).

[1]Elena M. Vella and Airlie Chapman are with the Department of Mechanical Engineering, The University of Melbourne, 3052 Parkville, Australia evella@student.unimelb.edu.au, airlie.chapman@unimelb.edu.au. [2]Nir Lipovetzky is with the School of Computing and Information Systems, The University of Melbourne, 3052 Parkville, Australia nir.lipovetzky@unimelb.edu.au

Digital Object Identifier (DOI): see top of this page.

the problem as a set-wise preference learning system and introduce a method to numerically represent preferences as vectors embedded in a feature space. We go on to introduce a preference model in §III and then extend the preference model to a group in §IV, where distinctiveness and group cohesion metrics are defined. An overview of the user study and results are discussed in §V, with concluding remarks in §VI.

II. Problem Formulation

We formulate a set-wise preference learning problem to study how a human user's preferences and more generally, the preferences of a population, are captured through a cost function from set-wise comparisons. We seek to design a control system that is robust to the subjective criteria of an individual and a population by accurately representing preferred preferences to meta-behaviors.

Notation: For $x \in \mathbb{R}^q$, we define the 1-norm of a vector $\|x\|_\mathbf{1} = \sum_{i=1}^{n} |x_i|$ and the 2-norm of a vector $\|x\|_\mathbf{2} = \sum_{i=1}^{n} |x_i|^2$. The indicator function $\mathbb{I}_A(\cdot)$ for a set $\mathcal{A} \in \mathbb{R}$ is defined as $\mathbb{I}_A(x) = 1$ if $x \in \mathcal{A}$ and otherwise equal to 0. The identity matrix I is a diagonal matrix with 1 on the dominant axis and 0 elsewhere. A pair of elements (x_i^1, x_i^2) appears uniquely in set $\mathcal{S}$, if and only if x_i^1 is preferred to x_i^2 for all possible set combinations $i = 1, ..., p$. A graph $\mathcal{G} = (V, S)$ is defined by a vertex set V of cardinality n and an ordered edge set $E \subseteq V \times V$ of cardinality m. If an edge exists from vertex v_j to vertex v_i it is expressed as $(v_j, v_i) \in E$. A directed acyclic graph is a graph with no directed cycles. We denote the standard normal cumulative density function as $\Phi(x)$. The function evaluates the probability that the value of a random variable $Y \sim \mathcal{N}(0, 1)$ is less than or equal to $x \in (-\infty, \infty)$. Similarly, the normal cumulative density function $\Phi_{\mu,\sigma}(x)$ describes the probability that the value of the random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ is less than or equal to x, denoted by $p(X \leq x)$. The cumulative density functions are related by $p(X \leq x) = \Phi_{\mu,\sigma}(x) = \Phi\left(x - \mu\right)/\sigma$; the inverse mapping is given by

$$\Phi^{-1}\left(p(X < x)\right) = \frac{x - \mu}{\sigma}. \tag{1}$$

A. Set-wise Preferences

The fundamental concept behind preference learning is a collection of empirical comparisons. In this work we question an individual, by comparing set-wise comparisons to meta-behaviors and performing a relative ranking, by a cost function, to find a user's preference profile.

Given a set of behaviors B, of size m. A set of behaviors $v = (w_1, w_2, \ldots, w_m) \subseteq B^m$ is termed a **meta-behavior**. Consider a set of meta-behaviors $\mathcal{B} \subseteq B^m$, we wish to collect a set of S_b **set-wise** comparison preferences $\mathcal{E}_\mathcal{B} = \{(v_i^1, v_i^2) | v_i^1, v_i^2 \in \mathcal{B}, i \in \{1, ..., m_b\}\}$, where m_b is the number of comparison preference elements. Preference comparisons are ranked relative to the response from a question, asking individuals to rank instances of meta-behaviors. An example question takes the form of "Comparing these meta-behaviors, which do you prefer more?"

Given each human has different preferences, we record a comparison set-wise preference ranking for each individual. We consider for the i^{th} comparison set, the preferred collection of meta-behaviors as $v_i^1 = (w_{i1}^1, w_{i2}^1, ..., w_{im}^1)$ ordered as the first set and the collection of non-preferred meta-behaviors as $v_i^2 = (w_{i1}^2, w_{i2}^2, ..., w_{im}^2)$ ordered as the second.

B. Set-wise Graph Interpretation

We may visualize an individual's set-wise comparison preferences using a preference graph. Consistent preferences may be depicted as acyclic graphs, yielding a partial order over preferences, while inconsistent preferences generate a cyclic graph.

For the k^{th} individual, the directed preference graph $\mathcal{G}_k = (V_k, \mathcal{E}_k)$ is defined by a vertex set $V_k \subseteq \mathcal{B}$ containing the compared collection of meta-behaviors $v^1, v^2 \in \mathcal{E}_\mathcal{B}$, and an ordered edge set $\mathcal{E}_k \subseteq V_k \times V_k \subseteq \mathcal{E}_\mathcal{B}$ indicating preferences among pairs of vertices. Here, a directed edge $(v_i, v_j) \in \mathcal{E}_k$ from vertex v_i to vertex v_j indicates the collection of meta-behaviors v_i is preferred to meta-behaviors v_j. Note there is at most one edge between each vertex in a pair, i.e. we do not consider self-contradictions. We assume that each preference is labeled as belonging to a given individual when constructing the individual's directed preference graph.

We extend this concept to an anonymous (*unlabeled*) group preference graph $\overline{\mathcal{G}} = (\overline{V}, \overline{\mathcal{E}}, W)$, in a situation where anonymized *unlabeled* preferences are collated from a general group of people, such that individual's preferences cannot be distinguished between. The graph $\overline{\mathcal{G}}$ has a vertex set $\overline{V} = \bigcup_{p=1}^{P} V_p$, the ordered set of edges $\overline{\mathcal{E}} \subseteq \mathcal{E}_\mathcal{B}$, and the associated edge weight set W, for the pth comparison preference, the total number of preferences $P = m_b \times q$. The group graph is formed by enumerating each set-wise preference, $a_{ij} = \sum_{p=1}^{P} \mathbb{I}_{\mathcal{E}_\mathcal{B}}\left((v_i, v_j)\right))$ for all $v_i, v_j \in \overline{V}$. The edge set $\overline{\mathcal{E}}$ contains the edge (v_i, v_j) if $a_{ij} - a_{ji} > 0$ (i.e. v_i is more preferred to v_j) or $a_{ij} = a_{ji} \neq 0$ and $i < j$ (i.e. v_i and v_j are equally preferred). The weights capture the preference variation of individuals in the group. The associated edge weight element of W is $w_{ij} = a_{ij}/(a_{ij} + a_{ji})$ for the comparison of meta-behaviors (v_i, v_j). Each vertex in the graph has a directed edge, the head is mapped to a preferred preference. The directed weighted edges describe the expectation of the preference ranking by the majority of individuals between vertices. In §III and §IV we will use the preference graph $\mathcal{G}$ or $\overline{\mathcal{G}}$ to extract a convex set of preferences for an ordinal optimization problem [13].

C. Human Comparator, Feature Embedding, and Classifiers

The subjectivity of an individual's preference is inevitable. To evaluate meta-behaviors effectively, we model each human's set-wise preferences to reflect an underlying cost function f_k. Fig. 1 is a basic input-output model, of the human comparator ranking meta-behaviors.

To address the challenge of human bias it is necessary to establish an empirical measurement for the cost function that is representative of human preferences, which can be generalized to find the best measure from the comparison set.

In an ordinal regression problem, the challenge arises of how to appropriately represent a cost function and generalize the expressed preferences for a given set of objectives preference data of an ordinal scale e.g "Strongly Agree", "Agree",... "Strongly Disagree". A collection of comparisons, on the other hand, develops a ranking cost function over the instances [14].

A key benefit of comparison studies over numerical or ordinal-ranking is individuals are less susceptible to ranking bias at the time of the event, this is known as batch effects in psychology [15]. For the remainder of the paper, we will refer to the preference learning technique for collections of comparisons, as preference learning. For a preference learning model, we may interpret and design the human as a comparator, a memoryless non-linear system. Such that when a human is given a set of behaviors, they output an ordered ranking of those behaviors.

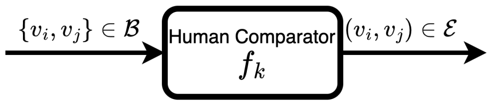


Fig. 1: Human modeled as a comparator, where f_k is the internal cost function.

A limiting factor to the approach of preference learning is we cannot infer preferences that are not in the set V_k of the preference graph. To overcome this challenge we propose to map each behavior to a point x in a feature space $\mathcal{X}$. There is an array of options for feature classification, to name a few support vector machine (SVM)-based classifier [16], extreme gradient boosting (XGBoost) classifier [17], and decision tree ensemble classifier [18]. Additionally, manual feature vector extraction for preference learning has been performed in [19].

In this work, we describe the compact feature vector as $x \in \mathcal{X}$ which has a mapping $h : B \to \mathcal{X}$ from the behavior set to a q-dimensional feature space $\mathcal{X} \subseteq \mathbb{R}^q$.

III. Developing Preference Profiles

To interpret individual set-wise comparisons meaningfully and effectively, we embed set-wise comparisons into a feature vector space. We amalgamate and build on concepts developed in [8] and [19] on preference learning. In the following sections, we propose a convex optimization set-wise preference learning problem with a global extremum. This allows us to successfully identify an optimally-preferred point for an individual from $\mathcal{G}_k$ and $h(\cdot)$ and later extend this notion to a group.

A. Cost Model For Preference Synthesis

Each element of a preference comparison pair $(v_i^1, v_i^2) \in \mathcal{E}_k$ where $v_i^1 = (w_{i1}^1, ..., w_{im}^1)$ and $v_i^2 = (w_{i1}^2, ..., w_{im}^2)$ is embedded and mapped by $x_i^1 = (h(w_{i1}^2), ..., h(w_{im}^2)) \in \mathcal{X}^m$ and $x_i^2 = (h(w_{i1}^2), ..., h(w_{im}^2)) \in \mathcal{X}^m$, respectively.

Set-wise preferences imply the existence of an underlying cost function $f_k : \mathcal{X}^m \to \mathbb{R}$ such that

$$f_k(x_i^1) \leq f_k(x_i^2) \Leftrightarrow (v_i^1, v_i^2) \in \mathcal{E}_k. \tag{2}$$

Restricting to length $m = 1$ meta-behaviors, i.e., a pairwise comparison preference, consider the quadratic cost function $f_k : \mathcal{X} \to \mathbb{R}$ proposed in [8] $f_k(x) = \|x - \overline{x}_k\|_2$, where $\overline{x}_k$ is a vector in $\mathcal{X}$ corresponding to an optimal preference for individual k.

We may estimate this function by considering the preference set as analogous to a set of affine classifications. In feature vector space, the preference set corresponds to a set of hyperplanes separating the pairs of behavior points with maximal distance to each point.

Firstly, we consider the i^{th} preference pair $(x_i^1, x_i^2) \in \mathcal{X} \times \mathcal{X}$, (2) is equivalent to the halfspace

$$g_i(x) = a_i^T x - b_i \leq 0 \ , \tag{3}$$

where $a_i = x_i^2 - x_i^1$ and $b_i = a_i^T (x_i^1 + x_i^2)/2$. The closed halfspace indicates that $g_i(x) \leq 0$ (the region of the feature space containing preferred behaviors) is convex but not affine. The two sets of points (preferred and non-preferred behaviors) can be linearly discriminated using a constant threshold. Geometrically this threshold corresponds to a hyperplane when $g(x) = 0$ formed by points yielding a constant-valued inner product with the vector a_i. The hyperplane divides $\mathcal{X}$ into two half-spaces, we interpret the two half-spaces as the preferred region $g(x) \leq 0$ and not preferred region $g(x) > 0$ of the space. In Fig. 2 the dashed lines correspond to the hyperplanes for the three preference pairs (x_{i1}^1, x_{i1}^2), (x_{i2}^1, x_{i2}^2) and (x_{i3}^1, x_{i3}^2), respectively.

We now extend the notion of a hyperplane created from a pairwise preference comparison to a set of preference comparisons. This is analogous to the human observing a task composed of meta-behaviors, instead of the individual observing individual behaviors that composed together make up the task.

Proposition 1: For the i^{th} meta-behavior preference pair $(x_i^1, x_i^2) \in \mathcal{X}^m \times \mathcal{X}^m$ where $x_i^1 = (x_{i1}^1, x_{i2}^2, ..., x_{im}^1)$ and $x_i^2 = (x_{i1}^2, x_{i2}^2, ..., x_{im}^2)$. For the set quadratic cost function $f_k : \mathcal{X}^m \to \mathbb{R}$ for individual k

$$f_k(x) = \sum_{j=1}^{m} \|x_j - \overline{x}_k\|_2^2 \tag{4}$$

then the preference inequality $f(x_i^1) \leq f(x_i^2)$ is equivalent to the halfspace constraint

$$g_i(x) = a_i^T x - b_i \leq 0 \ , \tag{5}$$

where $a_i = \sum_{j=1}^{m} \left(x_{ij}^2 - x_{ij}^1\right)$ *and* $b_i = \frac{1}{2} \sum_{j=1}^{m} \left(x_{ij}^2 - x_{ij}^1\right)^T \left(x_{ij}^1 + x_{ij}^2\right)$.

Proof: : The preference inequality is

$$\sum_{j=1}^{m} ||x_{ij}^1 - \overline{x}||_2^2 \leq \sum_{j=1}^{m} ||x_{ij}^2 - \overline{x}||_2^2$$

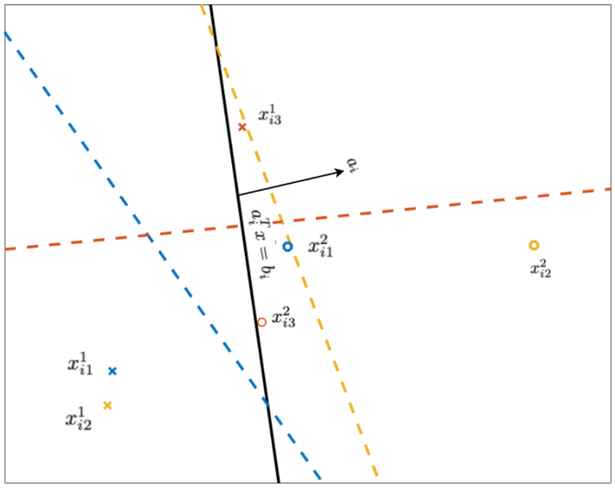


Fig. 2: The hyperplanes for three pairwise preference (x_{ij}^1, x_{ij}^2) for $j \in \{1, 2, 3\}$ (dashed lines). The hyperplane for a length three set-wise preference $x_i^1 = (x_{i1}^1, x_{i2}^1, x_{i3}^1)$ and $x_i^2 = (x_{i1}^2, x_{i2}^2, x_{i3}^2)$ (solid line).

with

$$\begin{aligned}
0 &\geq \sum_{j=1}^{m} (\|x_{ij}^1\|_2^2 - 2\overline{x}^T x_{ij}^1 + \|\overline{x}\|_2^2) \\
&\quad - (\|x_{ij}^2\|_2^2 - 2\overline{x}^T x_{ij}^2 + \|\overline{x}\|_2^2) \\
&= 2\overline{x}^T \sum_{j=1}^{m} (x_{ij}^2 - x_{ij}^1) + \sum_{j=1}^{m} (\|x_{ij}^1\|_2^2 - \|x_{ij}^2\|_2^2) \\
&= 2\overline{x}^T \sum_{j=1}^{m} (x_{ij}^2 - x_{ij}^1) - \sum_{j=1}^{m} (x_{ij}^2 - x_{ij}^1)^T (x_{ij}^1 + x_{ij}^2).
\end{aligned}$$

The result follows. ∎

Further, the set-wise halfspace (5) can be constructed using the pairwise halfspace (3). For the preference pair (x_{ij}^1, x_{ij}^2) if its halfspace (3) is $a_{ij}^T x - b_{ij} < 0$, the set-wise halfspace (5) has parameters $a_i = \sum_{j=1}^m a_{ij}$ and $b_i = \sum_{j=1}^m b_{ij}$.

The set-wise preference halfspace (5) for a length three meta-behavior preferences with $x_i^1 = (x_{i1}^1, x_{i2}^1, x_{i3}^1)$ and $x_i^2 = (x_{i1}^2, x_{i2}^2, x_{i3}^2)$ is depicted in Fig. 2.

B. Preference Polytope

Geometrically, the k^{th} individual's preference set can be described by a set of halfspaces (5) in feature vector space, with the i^{th} hyperspace associated with the sets of preferences $(v_i^1, v_i^2) \in \mathcal{E}_k$. Each preference comparison reduces the halfspace of $\mathcal{X}$, further constraining x. The intersection of the closed halfspaces defines a *preference polytope*, a region in feature vector space associated with the greatest preference. The intersection satisfies the system of linear inequalities created by the preferences and can be represented by the polytope $P_k = \{x \in \mathcal{X} | a_i^T x \leq b_i, \forall (v_i^1, v_i^2) \in \mathcal{E}_k\}$. An example preference polytope, is formed from the intersection of eight preference comparisons. The corresponding preferred halfspaces are given by the shaded interior region in Fig. 3.

The preference polytope can be unbounded for small $|\mathcal{E}_k|$ and poorly distributed preference comparisons. The preference polytope can also be empty for cyclic preference graphs and poorly selected embeddings.

The closed region of the polytope P_k can be used to determine preferred meta-behaviors; this process is often termed preference learning [3]. The polytope P_k can be built iteratively, with new set-wise preference comparisons presented to the individual over time. Given the polytope $P_k(t)$ at sample time t, the addition of the i^{th} preference at time $t+1$ forms the new polytope $P_k(t+1) = P_k(t) \bigcap \{x | g_i(x) \leq 0\}$. The strategic presentation of comparisons to the participant can rapidly reduce the volume of $P_k(t)$ over time.

C. Locating the extremum point - Chebyshev Center

As we have insufficient information within a bounded P_k to find $\overline{x}_k$ in (4), we may substitute an alternative point in $\mathcal{X}$. When presented with a bounded preference polytope P_k, we can define a *virtual point* x_c, that is maximally preferred within the polytope. We assert that the Chebyshev center is a suitable candidate for a hypothetical most-preferred preference point, as the Chebyshev center finds a globally optimal point in P_k. Feature vectors can then be compared to see which is closer to that point, which we denote as the *favored features* [19]. Finding the Chebyshev center is a convex optimization problem, hence we can find the global optimum of the affine set.

The Chebyshev center of P_k can be described geometrically as the center of the largest inscribed ball in P_k, also referred to as the in-center point. A visual interpretation of this is depicted in Fig. 3. Let the Chebyshev center x_c lie at the center of the largest possible ball $\mathcal{B}_r = \{x_c + u \,|\, \|u\|_2 \leq r\}$, inside P_k. We may obtain $\mathcal{B}_r$ by maximising r. For a weaker constraint, let $\mathcal{B}_r$ lie in the half-space $\|u\|_2 \leq r \implies a_i^T(x_c + u) \leq b_i$. The corresponding largest possible ball is given by $\sup\{a_i^T u \| \|u\|_2 \leq r\} = r\|a_i\|_2$. Therefore, the Chebyshev center lies within the P_k if and only if $a_i^T x_c + r\|a_i\|_2 \leq b_i, \forall (v_i^1, v_i^2) \in \mathcal{E}_k$. Given the ball radius $r \geq 0$, the Chebyshev center x_c can be found by solving the optimization

$$\begin{aligned}
(x_c, \overline{r}) &= \arg\max_{x,r} \; r, \\
&\text{s.t. } \; a_i^T x + r\|a_i\|_2 \leq b_i, \forall (v_i^1, v_i^2) \in \mathcal{E}_k.
\end{aligned} \tag{6}$$

The optimization is a linear program with many algorithms that can reliably and efficiently solve the problem [20]. The resulting x_c can then be used as a proxy to compare individuals' preferences.

IV. Group Preference Profiles

We extend the preference collection from an individual to a group. The preferences of a group of individuals are analyzed and a sparse optimization formulation is developed to introduce a distinctiveness metric that aggregates individuals with similar preference profiles and considers the effect of self-contradictory responses. We go on to consider anonymized unlabeled set-wise preferences, where we observe properties of the entire group's preferences. Assuming that the individual's maximal preference point

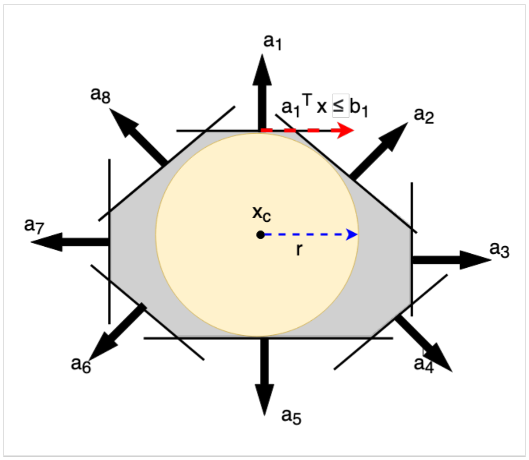


Fig. 3: The Chebyshev center x_c with radius r of the preference polytope P.

is drawn from a normal distribution, this provides further insight into the group's cohesion; via identification of the mode average most-preferred meta-behavior and minimum covariance bound.

A. Labeled Set-Wise Preferences

Aggregating individuals with similar preference profiles, from multiple individuals' respective preferences, we may devise a measure of the individuals' *distinctiveness*. The extremum preference value for individual k is denoted as $x + z^k$ where x is assumed to be a global reference preference measure and z^k the perturbation of individual k away from this reference. If the magnitude of z^k were small for all individuals, the group would exhibit similar preference behaviors. An individual with large z^k would, in turn, have distinctive preference behaviors compared to the group as a whole. Examining the preferences across each individual k, the selection problem for their i^{th} preference selection will generate the hyperspace $\left(a_i^k\right)^T \left(x + z^k\right) \leq b_i^k$. An optimization problem can then be posed to find x with small $\left\|z^k\right\|_1$ across all members using

$$(\overline{x}, \overline{z}) = \text{argmin}_{x, z=[z^1,\ldots,z^n]} \sum_{k=1}^{n} \left\|z^k\right\|_1, \quad (7)$$
$$\text{s.t. } \left(a_i^k\right)^T \left(x + z^k\right) \leq b_i^k, \forall (v_i^1, v_i^2) \in \mathcal{E}_k.$$

The accumulative 1-norm is used here to minimize z due to its sparsifying properties as it promotes sparse solutions with small or even zero $\left\|z^k\right\|_1$ [20]. When $\left\|z^k\right\|_1 = 0$, the reference preference measure x will satisfy all of the selection preferences for individual k. This subset of individuals will share a non-trivial intersection of their preference polytopes and an additional Chebyshev center selection could be performed to select a preference reference x_c with the characteristics described in §III-C.

B. Unlabeled Set-Wise Preferences

We seek to determine a metric to analyse the cohesion of the group, via a preference average and a covariance bounds

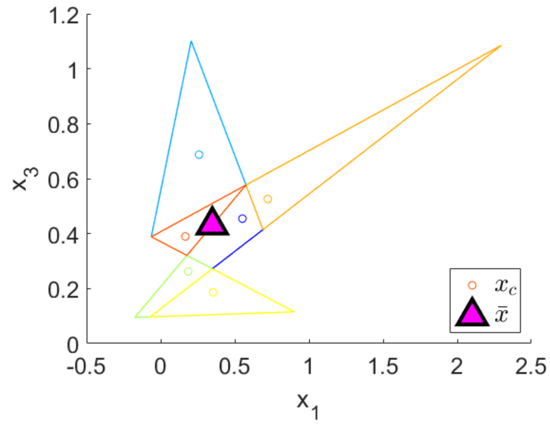


Fig. 4: Participant Chebyshev centers x_c compared with the aggregated preference optimum $\overline{x}$.

for a group's preference profiles; for the case where a preference may be expressed multiple times with contradictory responses.

Consider the selection of the perturbation z for each individual from a normal distribution $Z \sim \mathcal{N}(0, \Sigma)$, where Σ is a symmetric positive definite matrix. From (5), the i^{th} selection problem can subsequently be posed as a preferential selection of one choice over the other when the random variable $X_i = g_i(x+z) = a_i^T(x + z) - b_i$ is non-positive; the corresponding distribution is $\mathcal{N}(\mu = a_i^T x - b_i, \sigma^2 = a_i^T \Sigma a_i)$. By applying the inverse distribution mapping (1), the probability of a positive preference selection $p_i = p(X_i \leq 0)$ is then $-\Phi^{-1}(p_i)\sigma = \mu$. Assuming that the covariance of Z is bounded as $0 \preceq \Sigma \preceq \alpha^2 I$, then $\sigma \leq \alpha \left\|a_i\right\|_2$. For $p_i \geq 0.5$ then $\Phi^{-1}(p_i) \geq 0$ and

$$-\alpha \left\|a_i\right\|_2 \Phi^{-1}(p_i) \leq a_i^T x - b_i \leq 0. \quad (8)$$

Similarly, for $p_i < 0.5$ then

$$0 \leq a_i^T x - b_i \leq -\alpha \left\|a_i\right\|_2 \Phi^{-1}(p_i), \quad (9)$$

such that each preference constrains x to lie in what can be interpreted geometrically as a slab, i.e., a set of the form $\{x \in \mathbb{R}^q | \alpha \leq a^T x \leq \beta\}$ for scalars $\alpha \leq \beta$. With the data sampled from a finite population, the probability p_i is calculated based on a confidence interval $[p_i - \delta, p_i + \delta]$ projected onto the unit interval $[0, 1]$ as $\Delta = [\max(0, p_i - \delta), \min(1, p_i + \delta)]$. Here, 2δ is the width of the confidence interval band and is based on the margin of error calculated from a number of samples. Applying the central limit theorem for the binomial distribution is one approach to calculate the width with $\delta = Z\sqrt{1/4n_s}$ where Z is the Z-score associated with a confidence interval and n_s is the number of samples of the i^{th} preference [21].

We assume that $g_i(x)$ is constructed so that $X_i \leq 0$ for most of the group (i.e., $p_i \geq 0.5$), for example as per §II-B for $\overline{\mathcal{G}} = (\overline{V}, \overline{\mathcal{E}}, W)$ with $p_i = w_i \in W$. Using the constraints (8) and (9) over Δ, we may find the average preference measure $\overline{x}$ and minimum covariance bound $\overline{\alpha}$ for the group as

$$(\overline{x}, \overline{\alpha}) = \arg\min_{x,\alpha} \ \alpha \tag{10}$$
$$\text{s.t. } a_i^T x - \alpha \left\|a_i\right\|_2 \max(0, -\Phi^{-1}(p_i - \delta)) \leq b_i,$$
$$a_i^T x + \alpha \left\|a_i\right\|_2 \Phi^{-1}(\min(1, p_i + \delta)) \geq b_i,$$
$$\forall (v_i^1, v_i^2) \in \overline{\mathcal{E}}, p_i = w_i \in W.$$

For the i^{th} preference, the (positive) upper bound on the covariance α constrains the width of the slab containing x. Similarly, the closer p_i is to 0.5 (i.e. a split decision on the i^{th} preference among individuals), the narrower the slab.

V. User Study Human Preferences as a Trust Measure

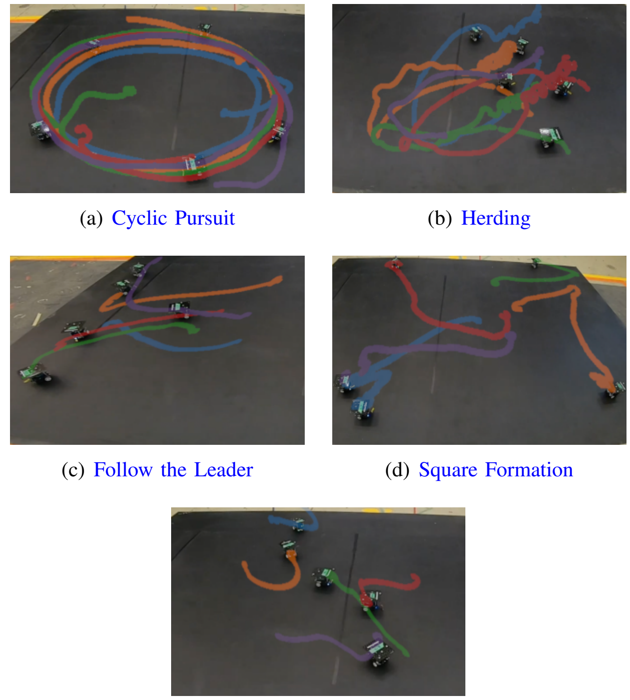
To understand the comparison of higher-dimensional objects, a user study experiment was performed by human observers. The working example focuses on swarm meta-behavior measured against a human's perception of what they are more trusting of.

There has been much discussion of the nature of human trust in robotic systems [22], [23] within a human-automation interaction paradigm. Trusted interactions emerge when humans delegate partial or full responsibility for a task to an autonomous system [24]. For a human to delegate autonomy to the robotic system, they must have sufficient trust in the system regarding proper task execution and safety, among other factors [25]. While previous investigations have focused predominantly on human interactions with industrial machines [23], in recent years a wider variety of collaborative human-robot interaction (HRI) scenarios have been investigated [25], [26]). Trust measures may be explicit or implicit, with the former type more objective and feature-specific (e.g. gaze, social cues, gestures). Implicit trust measures are more subjective, involving the beliefs and preconceptions of the individual in question [27]. Task-specific models that make use of either type of measure for absolute trust may not be able to robustly handle variations in how individuals express trust [26].

It is evident there is a gap in the flexibility of trust models and measures, to extend and develop the notion of human-robot trust it is required to have a stronger correlation for analysing trust. To address this, we consider human preferences and use preference learning techniques [11] for trust.

A user study experiment was conducted involving a swarm of ground-based robotic vehicles (unicycles) [28]. A unicycle model was designed to describe individual agent dynamics and a higher-level distributed multi-agent controller was employed to execute the desired swarm behaviors, to create a swarm meta-behaviors.

The executed swarm behaviors using the robotic platform for experimental purposes include: **1) cyclic pursuit**, where agents traverse a circle [29], **2) herding**, requiring agents to move from one location to another while maintaining collective cohesion [30], **3) leader following**, where a leader moves and is shadowed by all other agents [2], **4) square formation**, the agents distribute to the vertices of a square shape [2], **5) line formation**, agents evenly distribute to form a line [2]. The respective trajectories are depicted in Fig. 5. A total of forty-three participants participated in an online survey watching a collection of filmed swarm behaviors and recording their most preferred (trusted) meta-behaviors.



(a) Cyclic Pursuit (b) Herding (c) Follow the Leader (d) Square Formation (e) Line Formation

Fig. 5: Swarm behavior trajectory traces.

The notion of trust is nebulous, and potentially influenced by several characteristics. Any measure of trust will inherently prioritize certain characteristics and omit others, prescribing a model of trust that may not fully align with a participant's idiosyncratic evaluation of trust. Thus, the reliability of such a trust score representing the participant's absolute trust may not be certain. For these reasons, we have considered recording trust *preferences*. We have assumed that participants can infer and express their own trust preferences when presented with a pair of swarm behaviors. We have not considered cases when a participant cannot express a trust preference due to swarm behaviors having indistinguishable trustworthiness, or due to the participant rejecting the notion of trust as applicable to a swarm in general.

In the first section of the survey, participants were prompted to consider everyday devices they use and why they trust that item to perform its task reliably, safely, and consistently. It was our aim to prime the user with a self-association of trust that would help the participant answer the following questions [31]. It is imperative to highlight trust was not defined for each participant as it is a subjective matter we desired the individuals' interpretation of the meaning of trust. Participants watched a video of the unicycle swarm performing the behavior leader follower. A modified version of the *Trust Perception Scale-HRI* (TPS-HRI) [25], substituting the word *'swarm'* for *'robot'* fourteen questions relating to trust in the swarm were queried to the participant. The TPS-HRI trust survey was chosen because it covered the three aspects of automation trust (performance, process, purpose [32]) comprehensively.

In the second section of the survey, participants were presented with videos of the swarm meta-behaviors, preferred trust preferences were recorded for each individual when questioned *'Comparing the sets-wise swarm behaviors, which do you trust more?'*. Participants were not provided with an elucidation on the notion of trust, this was so that participants' trust preferences would be based exclusively on their observations of the swarm meta-behaviors seen in the videos and reduce batch effects.

A. Results

To visualize and analyze the preference consistency of each participant, the collected data from the online survey was used to create individual and a weighted group preference graph, illustrated in Fig. 8(a). Participants with contradictory responses represented as inconsistencies (cycles) in the graph were not considered.

The objective trust function was determined for each individual and the associated trust optimum (Chebyshev center) and respective preference polytope. In Fig. 4 we can observe the bounded preference polytopes of a subset of individuals and their associated x_c.[1] Individual x_c's have been aggregated to find the population trust optimum point $\overline{x}$ from (7). For a subset of twelve individuals, we form their individual preference polytopes and compare their optimum trust points (Chebyshev centers) x_c, (note some individuals have the same preference polytope, as they have the same responses to the survey).

We observe from the subset of participants, that the group trust extremum point, falls within the subset of individuals' optimum trust points, and is an appropriate group trust point candidate. The distinctiveness $\left\|z^k\right\|_1$ value and TPS-HRI trust score have been shown for the same subset of twelve individuals in Fig. 6. We observe there is a clear distinctiveness threshold value of $\left\|z^k\right\|_1 = 0.035$. Participants with a distinctiveness value $\left\|z^k\right\|_1 \leq 0.035$ fall within a trust score bounds of $[42\%, 56.5\%]$, may be interpreted to convey preferences accordant with the general group's preferences. In contradiction, participants with distinctiveness $\left\|z^k\right\|_1 > 0.035$ have preferences that oppose that of the general consensus of the group.

In Fig. 7 we extrapolate the notion of distinctiveness and trust bounds from Fig. 6 to the whole group. We may infer participants with a low distinctiveness value favor similar trust preferences to an average person in the group. Therefore we may generalize that participants with low distinctiveness would share similar preferences and be suitable to group together when selecting teams for human-swarm collaborative tasks.

A length one set-wise preference graph $\overline{\mathcal{G}} = (\overline{V}, \overline{\mathcal{E}}, W)$ illustrative of the partial ordering over swarm behaviors can be retrieved from the accumulation of unlabeled preference

[1]For visualization purposes only two dimensions of $\mathcal{X}$ are depicted.

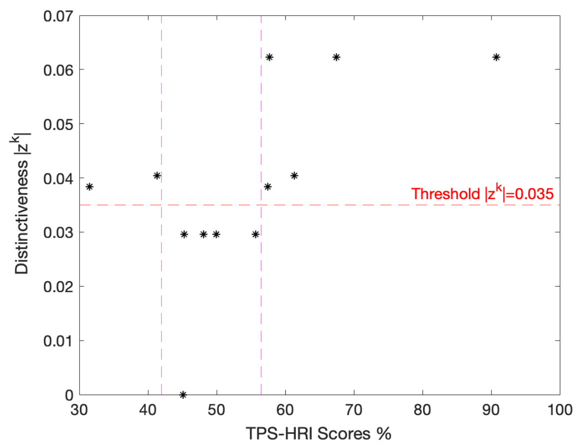


Fig. 6: Participant distinctiveness.

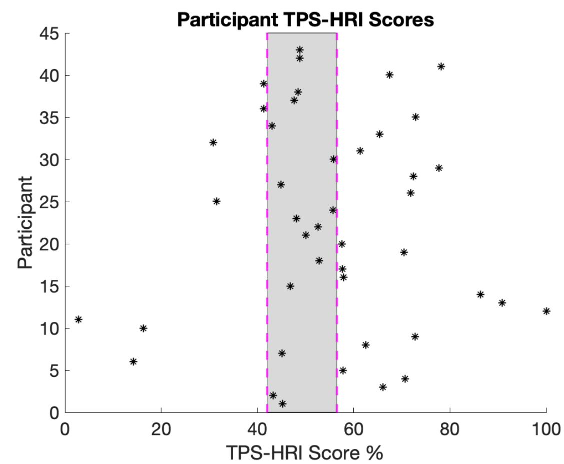
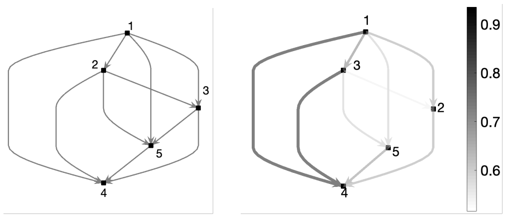


Fig. 7: Trust scores for participants as measured using TPS-HRI; the shaded region contains participants with distinctiveness $\left\|z^k\right\|_2 \leq 0.035$.

data. The edge-weightings from the population preference graph, may be used to derive the group's mean trust measure $\mu = \overline{x}$ and minimum covariance bound $\sigma \leq \overline{\alpha}$ from (10).

An evaluation of the performance of unlabeled length two set-wise preference data against the theoretical bounds is drawn in Table I. The mean μ is compared against the accumulated Chebyshev center $\overline{x}_c$ of the unweighted set-wise preference graph $(\overline{V}, \overline{\mathcal{E}})$. In Table II, of the individual participants, 54.05% have a Chebyshev center x_c that lies within the upper bound of one standard deviation α of the mean and 97.39% lie within two standard deviations 2α. The data correlates strongly with the theoretical bounds $p(-s < \left\|X - \mu\right\|_2 / \sigma < s) = \Phi(s) - \Phi(-s)$ for $s \in \{1, 2\}$. The unlabelled aggregated set-wise preference data is accordant with the presented group trust model in Fig. 8(b). The distance between the mean μ and the group's optimal trust point $\overline{x}$ is comparatively small, $\left\|\mu - \overline{x}_c\right\|_2 \leq 0.1\overline{\alpha}$. We may infer that the true value of $\overline{x}_k$ can be accurately depicted by the group set-wise preference graph $\overline{\mathcal{G}}$ and optimal solution of (10). We are able to therefore examine the cohesiveness of a group based on set-wise preference similarity.



(a) (b)

Fig. 8: a) A participant's acyclic set-wise preference graph; b) Weighted group set-wise preference graph (the gradient bar indicates the preference likelihood for the group).

TABLE I: Set-Wise Group Trust Statistical Information

Aggregate $\bar{\mathbf{x}}_{\mathbf{c}}$	**Mean, μ**	**Covariance Bound, $\bar{\alpha}$**
(-0.4130, 0.2203)	(-0.4166, 0.2217)	0.3203

TABLE II: Set-Wise Group Trust Preference Distribution

$\mathbf{s}$	$\mathbf{p}(-\mathbf{s} < \left\|\mathbf{x_c} - \mu\right\|_{\mathbf{2}} / \bar{\alpha} < \mathbf{s})$	$\Phi(s) - \Phi(-s)$
1	0.5405	0.6812
2	0.9739	0.9545

VI. Conclusions

In this work, we present a set-wise preference learning method for inferring the underlying cost structure to describe a group of human preference profiles. In particular, given a set of meta-behavior, we produce a partial order over the meta-behaviors, to infer the corresponding cost function. This work extends previous single-user preference models to group user models.

An example application has been provided for comparing swarm meta-behaviors. Individual and group preference models were verified through data collected from the online survey. A *distinctiveness* metric has been developed to describe the measure of an individuals set-wise preferences against a group of users. Accumulating and aggregating all unlabelled preferences from the online survey group, the weighted group preference graph was able to be created. The preference weights are applied to a sparse optimization problem to determine the group's *cohesion*. It is proposed that user profiles may be grouped by low distinctiveness and make the appropriate collective for complex cobotic tasks.

Future work includes exploring a sequential estimator that refines the estimate of the optimum point $\overline{x}$ as preferences are gathered over time. This would be coupled with the strategic selection of future preference pairs to present to the user, to optimise the exploration space.

Compliance With Ethical Standards

The authors declare that they have no conflict of interest. Informed consent has been obtained from all participants in accordance with the the Office of Research Ethics and Integrity at the University of Melbourne (reference #2021-22472-20996-4).

References

[1] T. Munzer, M. Toussaint, and M. Lopes, "Preference learning on the execution of collaborative human-robot tasks," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 879–885.

[2] P. Pierpaoli, T. T. Doan, J. Romberg, and M. Egerstedt, "A reinforcement learning framework for sequencing multi-robot behaviors," *arXiv e-prints*, pp. arXiv–1909, 2019.

[3] R. Herbrich, T. Graepel, P. Bollmann-Sdorra, and K. Obermayer, "Supervised learning of preference relations," *Proceedings des Fachgruppentreffens Maschinelles Lernen (FGML-98)*, pp. 43–47, 1998.

[4] W. Chu and S. S. Keerthi, "Support vector ordinal regression," *Neural computation*, vol. 19, no. 3, pp. 792–815, 2007.

[5] T. Graepel, K. Obermayer *et al.*, "Large margin rank boundaries for ordinal regression," in *Advances in large margin classifiers*. the MIT Press, 2000, pp. 115–132.

[6] W. W. Cohen, R. E. Schapire, and Y. Singer, "Learning to order things," *Advances in neural information processing systems*, vol. 10, p. 451, 1997.

[7] C. Cortes, M. Mohri, and A. Rastogi, "Magnitude-preserving ranking algorithms," in *Proceedings of the 24th international conference on Machine learning*, 2007, pp. 169–176.

[8] P. Kingston and M. Egerstedt, "Comparing apples and oranges through partial orders: An empirical approach," in *2009 American Control Conference*, Jun. 2009, pp. 5434–5439.

[9] R. Alami, R. Chatila, A. Clodic, S. Fleury, M. Herrb, V. Montreuil, and E. A. Sisbot, "Towards human-aware cognitive robots," in *The fifth international cognitive robotics workshop (the AAAI-06 workshop on cognitive robotics)*.

[10] K. Akash, W.-L. Hu, T. Reid, and N. Jain, "Dynamic modeling of trust in human-machine interactions," in *2017 American Control Conference (ACC)*, May 2017, pp. 1542–1548.

[11] J. Fürnkranz and E. Hüllermeier, "Preference learning and ranking by pairwise comparison," in *Preference learning*. Springer, 2010, pp. 65–82.

[12] E. M. Vella, D. A. Williams, A. Chapman, and C. Manzie, "Individual and team trust preferences for robotic swarm behaviors," in *2022 American Control Conference (ACC)*. IEEE, 2022, pp. 665–670.

[13] Y.-C. Ho, R. Sreenivas, and P. Vakili, "Ordinal optimization of deds," *Discrete event dynamic systems*, vol. 2, no. 1, pp. 61–88, 1992.

[14] C.-N. Fiechter and S. Rogers, "Learning subjective functions with large margins," in *ICML*, 2000, pp. 287–294.

[15] J. Díez, J. J. Del Coz, O. Luaces, and A. Bahamonde, "Clustering people according to their preference criteria," *Expert Systems with Applications*, vol. 34, no. 2, pp. 1274–1284, 2008.

[16] W. S. Noble, "What is a support vector machine?" *Nature biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006.

[17] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen *et al.*, "Xgboost: extreme gradient boosting," *R package version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.

[18] T. G. Dietterich, "Ensemble methods in machine learning," in *International workshop on multiple classifier systems*. Springer, 2000, pp. 1–15.

[19] P. Kingston, J. von Hinezmeyer, and M. Egerstedt, "Metric Preference Learning with Applications to Motion Imitation," in *Controls and Art*, A. LaViers and M. Egerstedt, Eds. Cham: Springer International Publishing, 2014, pp. 1–26.

[20] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[21] L. D. Brown, T. T. Cai, and A. DasGupta, "Interval Estimation for a Binomial Proportion," *Statistical Science*, vol. 16, no. 2, pp. 101 – 133, 2001.

[22] K. A. Hoff and M. Bashir, "Trust in automation: Integrating empirical evidence on factors that influence trust," *Human factors*, vol. 57, no. 3, pp. 407–434, 2015.

[23] J. D. Lee and K. A. See, "Trust in Automation: Designing for Appropriate Reliance," *Human Factors*, vol. 46, no. 1, pp. 50–80, Mar. 2004.

[24] S. Nahavandi, "Trusted Autonomy Between Humans and Robots: Toward Human-on-the-Loop in Robotics and Autonomous Systems," *IEEE Systems, Man, and Cybernetics Magazine*, vol. 3, no. 1, pp. 10–17, Jan. 2017, conference Name: IEEE Systems, Man, and Cybernetics Magazine.

[25] K. E. Schaefer, "Measuring Trust in Human Robot Interactions: Development of the "Trust Perception Scale-HRI"," in *Robust Intelligence and Trust in Autonomous Systems*, R. Mittu, D. Sofge, A. Wagner, and W. Lawless, Eds. Boston, MA: Springer US, 2016, pp. 191–218.

[26] B. C. Kok and H. Soh, "Trust in robots: Challenges and opportunities," *Current Robotics Reports*, vol. 1, no. 4, pp. 297–309, 2020.

[27] R. E. Yagoda and D. J. Gillan, "You want me to trust a robot? the development of a human–robot interaction trust scale," *International Journal of Social Robotics*, vol. 4, no. 3, pp. 235–248, 2012.
[28] E. Schoof, C. Manzie, I. Shames, A. Chapman, and D. Oetomo, "An experimental platform for heterogeneous multi-vehicle missions," in *Proceedings of the International Conference on Science and Innovation for Land Power, Adelaide, Australia*, 2018, pp. 5–6.
[29] J. Marshall, M. Broucke, and B. Francis, "Formations of vehicles in cyclic pursuit," *IEEE Transactions on Automatic Control*, vol. 49, no. 11, pp. 1963–1974, Nov. 2004.
[30] A. Pierson and M. Schwager, "Bio-inspired non-cooperative multi-robot herding," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, May 2015, pp. 1843–1849.
[31] E. U. Weber and E. J. Johnson, "Constructing preferences from memory," *The Construction of Preference, Lichtenstein, S. & Slovic, P.,(eds.)*, pp. 397–410, 2006.
[32] J. D. Lee and N. Moray, "Trust, control strategies and allocation of function in human-machine systems," *Ergonomics*, vol. 35, no. 10, pp. 1243–1270, Oct. 1992.