Using Exploratory Search to Learn Representations for Human Preferences

Nathaniel Dennler
dennler@usc.edu
University of Southern California
United States of America

Stefanos Nikolaidis
nikolaid@usc.edu
University of Southern California
United States of America

Maja Matarić
mataric@usc.edu
University of Southern California
United States of America

ABSTRACT

Robots that interact with humans must adapt to the different preferences of human users. However, the time and effort needed for non-expert users to specify their preferences for a robot are a barrier to effective robot adaptation. Better representations of user preferences in the form of learned features have the potential to facilitate robot adaptation. In this work, we propose a method to learn representations using Contrastive Learning from Exploratory Actions (CLEA) that leverages data automatically collected from an interactive signal design process to better learn user preferences. We show that using data collected automatically from the design process can aid with learning user preferences compared to purely self-supervised learning.

CCS CONCEPTS

• **Human-centered computing** → **HCI design and evaluation methods**; • **Computing methodologies** → *Semi-supervised learning settings*; **Learning from implicit feedback**.

KEYWORDS

Preference Learning; Exploratory Search; Human-Robot Interaction.

ACM Reference Format:
Nathaniel Dennler, Stefanos Nikolaidis, and Maja Matarić. 2024. Using Exploratory Search to Learn Representations for Human Preferences. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24 Companion), March 11–14, 2024, Boulder, CO, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3610978.3640745

1 INTRODUCTION

In-home robots are expected to interact and collaborate with humans in a wide variety of social contexts, home environments, and tasks. Exactly how a given robot should perform in these varying contexts is unclear before the robot is deployed and cannot be deciphered from the environment alone [25, 11, 10]. A promising approach to enabling robots to adapt to unique contexts is to allow users to specify the robot's behavior themselves.

However, teaching a robot is a non-trivial task for any user [1] and the effort required to correctly communicate preferences to a robot can be a barrier to using the robot [17, 18]. A natural way for users to specify robot behavior is to select their favorite behavior from a small set of options [23, 8, 15, 14]. In using this approach, however, the user is constrained to seeing only a very small subset of possible behaviors the robot is capable of, making it difficult for the user to understand the robot's capabilities.

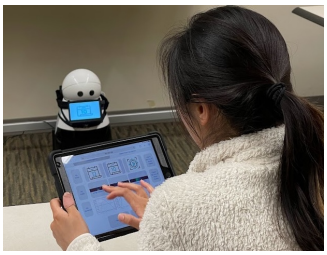


Figure 1: Participant engaging in the signal design task. We leverage the user's exploratory actions during the design process to better learn representations for signaling preferences for all future users of the system.

We take inspiration from *exploratory search* used in human-computer interaction (HCI), where participants spend time searching through a vast array of information to both learn about a topic and to determine exactly what they are looking for [19, 6]. Exploratory search is directly applicable to human-robot interaction (HRI) because users often have to develop mental models of a robot's true capabilities through seeing the robot in action [22, 26]. Information collected from the signal design process (shown in Figure 1) can aid preference learning. Users take data-generating actions in search interfaces to explore, filter, and examine different search results. The goal of our work is to evaluate how we can leverage exploratory data automatically collected from exploratory actions to learn representations useful for preference learning.

Existing approaches to learn representations for preference learning are to use self-supervised methods to encode robot behaviors or collect explicit similarity data [5]. Self-supervised representations encode information to reconstruct robot behavior, but may not be useful or relevant for human preferences. Collecting explicit similarity data may also be burdensome to the user. We propose Contrastive Learning from Exploratory Actions (CLEA) as a method to supplement self-supervised learning with data generated from exploratory search actions, shown in Figure 2. We show that CLEA can learn better representations than self-supervised learning alone, toward eliciting preferences from non-expert users in the context of a signal design task.

2 BACKGROUND

Exploratory Search Exploratory search is a concept from human-computer interaction (HCI) that describes the process through

HRI '24 Companion, March 11–14, 2024, Boulder, CO, USA

ACM ISBN 979-8-4007-0323-2/24/03
https://doi.org/10.1145/3610978.3640745

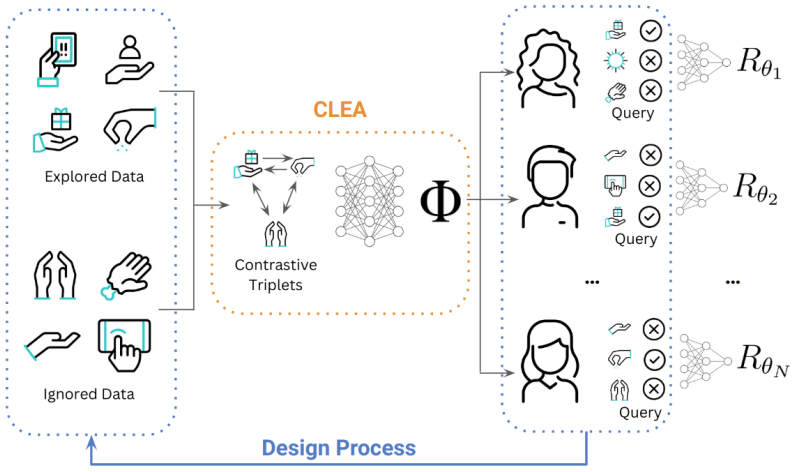


Figure 2: Overview of the proposed framework. Our method, Contrastive Learning from Exploratory Actions (CLEA), optimizes a contrastive objective using data gained from design processes to learn trajectory features that can be used to elicit preferences for individual users.

which users interact with data systems to search for a goal. For example, when a person wants to find a restaurant in an unfamiliar city, they must jointly understand what options are available to them and additionally consider their own preferences for restaurants in general. In exploratory search, the exact goal is unknown ahead of time because the user is unfamiliar with the search topic and how the goal can be achieved [19]. The contrasting search paradigm is information retrieval [24], where the user knows exactly what they need to find. Works in HCI have described several interfaces that facilitate exploratory search in different contexts: children reading about animals [2], users searching for restaurants in new cities [12], and students looking for research advisors [21]. The main focus of these works was learning how to represent the different kinds of data users may come across to facilitate the search process [16]. In our work, we apply these ideas from exploratory search to the process of eliciting user preferences in the context of robot signaling.

Eliciting User Preferences for Robot Trajectories Research in eliciting user preferences in robotics has focused on using different modalities for users to specify their preferences. For example, users can provide demonstrations [20], choose between example behaviors [23], provide rankings of example behaviors [9], or provide corrections as feedback [4]. These methods typically use representations of the different actions robots can take to elicit behavior preferences from a user. The features are often hand-coded, but it can be difficult to know ahead of time what aspects of robot behavior users care about. Previous work has proposed learning representation by querying users about which two robot actions are most similar out of three actions [5], however, this requires an added data collection step. In our work, we propose to leverage the data from exploratory search that users generate automatically to learn representation of robot behaviors. The data from exploratory actions can be combined with similarity queries or other human-generated information to learn better representations of robot behaviors that facilitate later elicitation of preferences.

3 APPROACH

Our approach aims to address a key problem in eliciting user preferences: the challenge associated with collecting data for learning representations. We leverage data collected automatically from exploratory search actions taken by the user during the design process. In particular, we use the information collected from the search-based interface (Figure 3b) to learn representations. We use the preference information collected from the query-based interface (Fig. 3a) to evaluate the effectiveness of representations.

Signal Design Task We aim to learn representations for eliciting preferences using data collected from the signal design study developed in our previous work [13]. The design study participants were tasked with designing four signals for a robot that engaged the user in an item-finding task: an idling signal, a searching signal, a "have item" signal, and a "have information" signal. For each signal, the user selected three different signal components to be played on the robot: a video component played on the robot's screen, an audio component played through the robot's speaker, and a movement component played on the robot's head. Users selected these components by using the two-interface application shown in Figure 3. The query-based interface (Fig. 3a) allows users to specify their favorite signal component from a set of three candidates. The search-based interface (Fig. 3b) allows users to enter search terms and scroll through multiple options to select their favorite signal component. Participants were free to use either interface to design their signal. Once participants finished designing one of the four signals, they pressed "submit" to move to the next signal. The full descriptions of the design process, participant demographics, and study details are provided in our previous work [13].

Learning Representations from Exploratory Actions. Our approach for learning trajectory representations relies on the insight that users tend to select options that they are seriously considering in exploratory search [3]. Where self-supervised learning methods learn representations of trajectories that are functionally

similar, people consider *semantically* similar items as more similar– a head motion moving side to side with a neutral expression is very similar to a side to side motion with a positive expression in the space of trajectories, but their semantic interpretations (i.e., fear vs. excitement) are vastly different. We aim to learn these distinctions by using a contrastive objective to learn representations that are useful for understanding user preferences.

We created a dataset of exploratory actions exhibited by our study participants. We denoted a search result as a set of k generated trajectories $S \in \Xi^k$. Each search result can generate an arbitrary number of relevant trajectories. We partitioned S into the set of trajectories that the user chose to explore in detail (by playing them on the physical robot), $S_{explored}$ and all other trajectories in the search result, $S_{ignored} = S \backslash S_{explored}$. We generated a triplet (ξ_a, ξ_p, ξ_n) by sampling the anchor trajectory ξ_a and the positive trajectory ξ_p from $S_{explored}$ and the negative trajectory ξ_n from $S_{ignored}$. Triplets were also generated by swapping the sampled sets $S_{explored}$ and $S_{ignored}$. To train an embedding network to generate a representation consisting of d-dimensional features $\phi : \Xi \rightarrow \mathbb{R}^d$, we used the triplet loss:

$$\mathcal{L}_{trip.}(\xi_a, \xi_p, \xi_n) = \max\left[d(\xi_a, \xi_p) - d(\xi_a, \xi_n) + \alpha, 0\right] \quad (1)$$

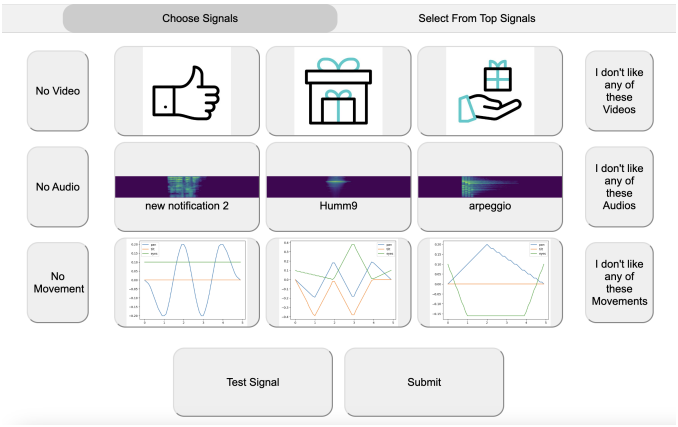
where the distance function, d, is defined as the Euclidean distance between features of the trajectory, $d(\xi_1, \xi_2) = ||\phi(\xi_1) - \phi(\xi_2)||_2^2$, and $\alpha \geq 0$ represents the minimum desired distance between trajectory pairs. Since either trajectory could be the anchor or positive pair, we made the final contrastive loss symmetric by:

$$\mathcal{L}_{cont.}(\phi, \xi_a, \xi_p, \xi_n) = \mathcal{L}_{trip.}(\phi, \xi_a, \xi_p, \xi_n) + \mathcal{L}_{trip.}(\phi, \xi_p, \xi_a, \xi_n) \quad (2)$$

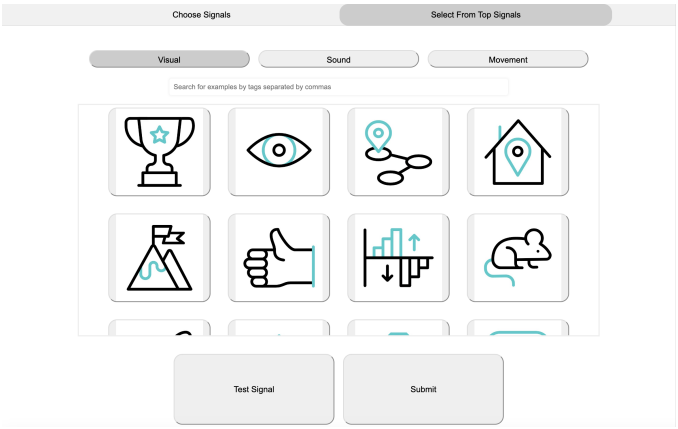
This loss can also be combined with reconstruction losses to jointly learn representations that can both reconstruct trajectories and consider semantically similar trajectories for each signal design task. The data collected in our study for learning representations contains 520 search actions, with an average of 2.673 explored options and 24.506 unexplored options for each search action across 25 participants.

Eliciting Preferences from User Choices Preference learning provides a framework for inferring how a user would like to act. In this work, we consider users as having preferences over trajectories, $\xi \in \Xi$, where a user's preferences are modeled by a reward function, $R : \Xi \rightarrow \mathbb{R}$, that cannot directly be evaluated; however, the reward function can influence how users choose between different example robot behaviors. We aimed to estimate R with a neural network R_θ using data collected from user choices. We trained this network by making explicit queries of the user by asking them to choose the best trajectory from a set of k trajectories, $\{\xi_1, \xi_2, ..., \xi_k\}$ that we call $Q \in \Xi^k$. In our study, we used $k = 4$. We adopt the Bradley Terry model of rational choice [7] for calculating the probability of choosing a specific trajectory from the set Q:

$$P(\xi_i | Q, \theta) = \frac{e^{R_\theta(\phi(\xi_i))}}{\sum_{\xi_k \in Q} e^{R_\theta(\phi(\xi_k))}} \quad (3)$$



(a) Query-based interface for choosing among three signals per modality.



(b) Search-based interface for browsing all options for each modality.

Figure 3: Design interface for the robot signal design tool. Participants chose freely between query-based interfaces and search-based interfaces to design signals.

We maximize this probability from the query data we collected in our study using a cross-entropy loss as in several previous works for learning preferences with neural networks [5]. In practice, we allowed users to specify that none of the trajectories are fit for the signal that they are designing. We denoted this option as $\xi_\emptyset$, and defined its reward as $R_\theta(\phi(\xi_\emptyset)) = 0$. For evaluation, we collected 1035 query preference responses from 25 participants.

4 EXPERIMENTS

Baselines We compared 5 methods for learning representations for trajectories: (1) ***Random***, a randomly-initialized network to extract non-linear features; (2) ***Autoencoder (AE)***, a network trained with the objective of reconstructing the original trajectory using a mean-squared error loss; (3) ***Variational Autoencoder (VAE)***, a network trained with the same reconstruction as the AE and a variational term that encourages the latent space to follow a normal distribution; (4) ***CLEA***, a network trained using the loss function $\mathcal{L}_{contrastive}$ as described in Section 3; and (5) ***CLEA+AE***, a network trained on the combined reconstruction loss and contrastive loss.

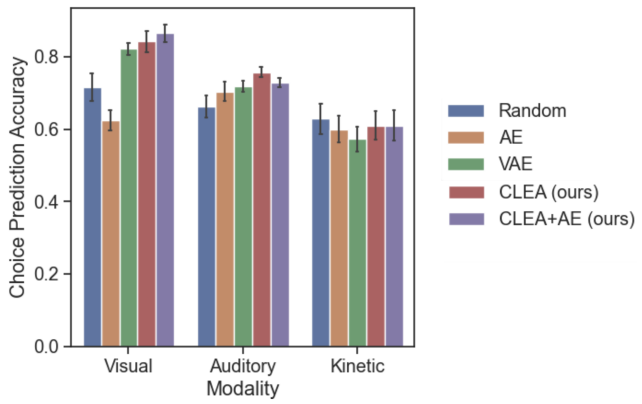


Figure 4: Overall choice prediction accuracy by modality and approach. Error bars represent the standard error of the mean accuracy, aggregated over all four tasks. Using a contrastive objective from exploratory actions can help increase prediction accuracy in downstream preference learning tasks.

Modality	Method	Idle	Search	Item	Info
Visual	Random	.81 ± .07	.63 ± .04	.85 ± .00	.57 ± .00
	VAE	.86 ± .02	**.73 ± .00**	.86 ± .01	.84 ± .02
	AE	.73 ± .01	.48 ± .01	.62 ± .03	.66 ± .03
	CLEA (ours)	**.92 ± .02**	.69 ± .01	.89 ± .00	.87 ± .03
	CLEA+AE (ours)	**.92 ± .01**	**.73 ± .01**	**.92 ± .00**	**.89 ± .00**
Auditory	Random	.73 ± .00	.71 ± .00	.72 ± .03	.49 ± .00
	VAE	.75 ± .01	.73 ± .01	.76 ± .00	.63 ± .02
	AE	.77 ± .01	.70 ± .02	**.78 ± .00**	.57 ± .00
	CLEA (ours)	**.82 ± .02**	**.75 ± .01**	.76 ± .02	**.69 ± .03**
	CLEA+AE (ours)	.75 ± .01	.73 ± .00	.77 ± .01	.66 ± .03
Kinetic	Random	.44 ± .03	**.73 ± .02**	**.56 ± .00**	**.79 ± .02**
	VAE	.45 ± .00	.58 ± .04	.51 ± .03	.75 ± .05
	AE	.45 ± .00	.67 ± .02	.52 ± .02	.76 ± .04
	CLEA (ours)	.44 ± .03	.70 ± .02	.54 ± .03	.76 ± .00
	CLEA+AE (ours)	**.48 ± .03**	**.73 ± .02**	.46 ± .02	.76 ± .00

Table 1: Predicting user preference choices from the query interface. The highest prediction accuracy is shown in bold and the second highest in blue.

Evaluation To evaluate these methods of learning representations, we collected data on the choices users made when presented with preference queries. Users could select one signal component from a set of three candidates in the query-based interface (shown in Figure 3a), or say that they did not like any of the options. We learned a separate reward network for each of the four signals (idle, search, has item, and has info) for each participant and each modality that predicts which of the four choices the user makes.

5 RESULTS

We evaluated our method using a leave-one-out cross-validation evaluation setting. We first trained a model for each modality to learn representations from exploratory search data using each of the approaches described in Section 3. For the visual modality, we used a still image of the video and learned representations with a convolutional neural network. For the auditory modality, we used the time-frequency spectrogram, and learned representations with a convolutional neural network. For the kinetic modality, we used the sequence of joint states and learned representations using a recurrent neural network. We used identical networks to generate the embedding for each method. We learned representations that were 128-dimensional vectors, similar in size to other works that learn representations for preference learning [5].

After training the representation networks, we trained models to learn the user's reward function with query data from these representations. We used a feed-forward network for all modalities and learned a separate reward model for each modality, participant, and signal type. We did this across 5 random seeds to compare the different approaches.

We present the results aggregated across the four types of signals in Figure 4. We found that in the visual modality, CLEA+AE performed the best, for the auditory modality, CLEA performed the best, and for the kinetic modality, all approaches performed approximately equally well.

We provide de-aggregated signal type results in Table 1. We found that using a contrastive loss either with CLEA or CLEA+AE resulted in the highest accuracy for nine of the twelve signals across all the modalities. For the three tasks that the contrastive loss was not used in the approach with the highest accuracy, it was used in the approach with the second-highest accuracy.

6 DISCUSSION AND FUTURE WORK

We found that using a contrastive loss is beneficial for downstream preference learning tasks – preferences exhibited by different users across different signals. We hope to reduce the number of models needed for learning representations by conditioning on the task in future work. We found the most success with using the contrastive loss in visual and auditory signal preferences. A possible reason for this finding is that exploratory search in robot learning requires a way to briefly summarize robot behaviors so that users can quickly evaluate what each behavior represents. Our study participants were easily able to understand the representation of the video, which is inherently visual. Participants had more difficulty understanding the spectrograms that represented auditory signals, but were familiar with viewing sounds as waveforms, and drew connection to that representation when reviewing the spectrograms. Users had the most difficulty understanding how the graphs of joint values translated to actual motion. We hypothesize that the visual descriptions of these signals are important for both helping the user with accurately performing exploratory actions and for eliciting preferences from these exploratory actions.

Future work can investigate how to generate these descriptions of behavior. We can also investigate how to leverage exploratory search data more; we hypothesize that sharing data between similar tasks can help to facilitate the learning process, and augmenting user data with data from similar users can also increase the efficiency of preference learning.

We showed that by incorporating the information gained from the user's exploratory actions during a signal design process, we can learn representations that are useful for downstream preference learning tasks. This shows the importance of interaction design in preference learning interfaces, which are used to both design signals and collect data for learning representations.

REFERENCES

[1] Baris Akgun, Maya Cakmak, Jae Wook Yoo, and Andrea Lockerd Thomaz. 2012. Trajectories and keyframes for kinesthetic teaching: a human-robot interaction perspective. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, 391–398.

[2] Garrett Allen, Benjamin L Peterson, Dhanush Kumar Ratakonda, Mostofa Najmus Sakib, Jerry Alan Fails, Casey Kennington, Katherine Landau Wright, and Maria Soledad Pera. 2021. Engage!: co-designing search engine result pages to foster interactions. In *Interaction Design and Children*, 583–587.

[3] Kumaripaba Athukorala, Dorota Głowacka, Giulio Jacucci, Antti Oulasvirta, and Jilles Vreeken. 2016. Is exploratory search different? a comparison of information search behavior for exploratory and lookup tasks. *Journal of the Association for Information Science and Technology*, 67, 11, 2635–2651.

[4] Andrea Bajcsy, Dylan P Losey, Marcia K O'Malley, and Anca D Dragan. 2018. Learning from physical human corrections, one feature at a time. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 141–149.

[5] Andreea Bobu, Yi Liu, Rohin Shah, Daniel S Brown, and Anca D Dragan. 2023. Sirl: similarity-based implicit representation learning. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 565–574.

[6] Oliver Bown, Sam Ferguson, Liam Bray, Angelo Fraietta, and Lian Loke. 2019. Facilitating creative exploratory search with multiple networked audio devices using happybrackets. In *New Interfaces for Musical Expression*. NIME.

[7] Ralph Allan Bradley and Milton E Terry. 1952. Rank analysis of incomplete block designs: i. the method of paired comparisons. *Biometrika*, 39, 3/4, 324–345.

[8] Daniel Brown, Russell Coleman, Ravi Srinivasan, and Scott Niekum. 2020. Safe imitation learning via fast bayesian reward inference from preferences. In *International Conference on Machine Learning*. PMLR, 1165–1177.

[9] Daniel Brown, Wonjoon Goo, Prabhat Nagarajan, and Scott Niekum. 2019. Extrapolating beyond suboptimal demonstrations via inverse reinforcement learning from observations. In *International conference on machine learning*. PMLR, 783–792.

[10] Barbara Bruno et al. 2019. The caresses eu-japan project: making assistive robots culturally competent. In *Ambient Assisted Living: Italian Forum 2017 8*. Springer, 151–169.

[11] Fabio Maria Carlucci, Lorenzo Nardi, Luca Iocchi, and Daniele Nardi. 2015. Explicit representation of social norms for social robots. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 4191–4196.

[12] Joseph Chee Chang, Nathan Hahn, Adam Perer, and Aniket Kittur. 2019. Searchlens: composing and capturing complex user interests for exploratory search. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 498–509.

[13] Nathaniel Dennler, David Delgado, Daniel Zeng, Stefanos Nikolaidis, and Maja J Matarić. 2023. The rosid tool: empowering users to design multimodal signals for human-robot collaboration. In *Experimental Robotics: The 18th International Symposium*. Springer.

[14] Johannes Fürnkranz, Eyke Hüllermeier, Weiwei Cheng, and Sang-Hyeun Park. 2012. Preference-based reinforcement learning: a formal framework and a policy iteration algorithm. *Machine learning*, 89, 123–156.

[15] Hong Jun Jeon, Smitha Milli, and Anca Dragan. 2020. Reward-rational (implicit) choice: a unifying formalism for reward learning. *Advances in Neural Information Processing Systems*, 33, 4415–4426.

[16] Tingting Jiang. 2014. Exploratory search: a critical analysis of the theoretical foundations, system features, and research trends. *Library and information sciences: Trends and research*, 79–103.

[17] William R King and Jun He. 2006. A meta-analysis of the technology acceptance model. *Information & management*, 43, 6, 740–755.

[18] Paul Legris, John Ingham, and Pierre Collerette. 2003. Why do people use information technology? a critical review of the technology acceptance model. *Information & management*, 40, 3, 191–204.

[19] Gary Marchionini. 2006. Exploratory search: from finding to understanding. *Communications of the ACM*, 49, 4, 41–46.

[20] Andrew Y Ng, Stuart Russell, et al. 2000. Algorithms for inverse reinforcement learning. In *Icml*. Vol. 1, 2.

[21] Behnam Rahdari, Peter Brusilovsky, Dmitriy Babichenko, Eliza Beth Littleton, Ravi Patel, Jaime Fawcett, and Zara Blum. 2020. Grapevine: a profile-based exploratory search and recommendation system for finding research advisors. *Proceedings of the Association for Information Science and Technology*, 57, 1, e271.

[22] Matthew Rueben, Jeffrey Klow, Madelyn Duer, Eric Zimmerman, Jennifer Piacentini, Madison Browning, Frank J Bernieri, Cindy M Grimm, and William D Smart. 2021. Mental models of a mobile shoe rack: exploratory findings from a long-term in-the-wild study. *ACM Transactions on Human-Robot Interaction (THRI)*, 10, 2, 1–36.

[23] Dorsa Sadigh, Anca D Dragan, Shankar Sastry, and Sanjit A Seshia. 2017. *Active preference-based learning of reward functions*.

[24] Mark Sanderson and W Bruce Croft. 2012. The history of information retrieval research. *Proceedings of the IEEE*, 100, Special Centennial Issue, 1444–1451.

[25] Sarah Sebo, Brett Stoll, Brian Scassellati, and Malte F Jung. 2020. Robots in groups and teams: a literature review. *Proceedings of the ACM on Human-Computer Interaction*, 4, CSCW2, 1–36.

[26] Aaquib Tabrez, Matthew B Luebbers, and Bradley Hayes. 2020. A survey of mental modeling techniques in human–robot teaming. *Current Robotics Reports*, 1, 259–267.