Singing the Body Electric: The Impact of Robot Embodiment on User Expectations

Author Names Omitted for Anonymous Review.

***Abstract*—Users develop mental models of robots to conceptualize what kind of interactions they can have with the robot. These conceptualizations are often formed before interaction with the robot and are based on its physical design and embodiment. We propose to use multimodal features of robot embodiments to predict what kinds of expectations users will have about robot social and physical capabilities. We show that using these features can provide information about mental models of users' expectations of robots.**

I. Introduction

When interacting with novel systems, users develop mental models of how they can interact with the systems. These mental models are instrumental to allow users to naturally interact with systems that can be arbitrarily complex technically [24]. Human-computer interaction has successfully used the concept of design metaphors to develop visual interfaces and interactions that are easy for users to learn [25, 13, 15, 18].

Understanding how users perceive robot expectations is difficult because robots are physically embodied, and have a wider variety of form factors and ways to interact with people than computers. These additional modes of communication contribute to robots' increased social presence [8]. In this work, we propose leveraging information about how the robot is physically embodied to understand how people form mental models of the robot. We use the Metaphors to Understand Functional and Social Anticipated Affordances (MUFaSAA) Dataset [10] that contains 165 robot embodiments and their associated ratings of social and functional attributes. Our preliminary results show that including information about how robots are physically embodied can provide information about how people will expect that robot to behave socially and functionally, informing how interactions with specific robots should be designed.

II. Background

This section provides a brief overview of the concepts of embodiment in robotics and mental models of robots.

A. Robot Embodiment

Robots are unique from computer-based agents because they have the ability to interact with and manipulate the physical world. Due to this interactive nature, robots can provide a stronger social presence because they can leverage additional modes of communication inaccessible to other forms of technology, such as proxemics, gaze, and gestures [8]. Previous work in robotics and psychology has shown that people form expectations from initial observations of new technology before extensive use [11, 4, 21, 19]. Thus the embodiment, i.e., the physical design of the robot, is a key component to understanding how users form expectations of how robots will act in day-to-day use. For robots to be effective, they must understand the social and functional expectations that users place on them so that they can perform to these expectations. A failure to meet these expectations negatively impacts adoption of these systems [6, 7, 19].

B. Mental Models and Design Metaphors

Mental models are conceptual frameworks that people develop to understand how they can interact with other agents [17]. Previous work has shown that users with mental models that accurately represent the underlying, more complex system are more effective at using the system [16]. Mental models are often based around what capabilities robots are expected to perform [24]. These models are formed before interaction, but are updated as users learn more about how systems work. However, even after interaction users can still form incorrect mental models of a robot's true capability. For example, one study showed that robots using speech are expected to better at physical manipulation despite the fact that these two capabilities are unrelated technically [6]. Understanding the initial mental models that users form around the robot are important to correctly calibrate a robot's true capabilities to avoid misleading users' expectations.

A tool that is often used to set expectations of new technology are design metaphors. Design metaphors associate unfamiliar systems with familiar and related concepts to provide a user with schemas to interact with novel systems. For example, in one study chatbots were described with different design metaphors (e.g., "a toddler", "a trained professional", "an inexperienced teenager", etc.), shaping user perceptions of the chatbot's warmth and competence, thereby affecting both the users' pre-interaction intention to use the system and their subsequent intention to adopt the system post-interaction [15]. By understanding the metaphors people use to understand the embodiment, we can gain information about how they expect to interact with it.

III. Methods

Our work leverages the MUFaSAA dataset to predict the six constructs that describe the social and functional expectations users place on the robots. These constructs represent the users' mental models of the robot. We outline our process for predicting these expectations below.

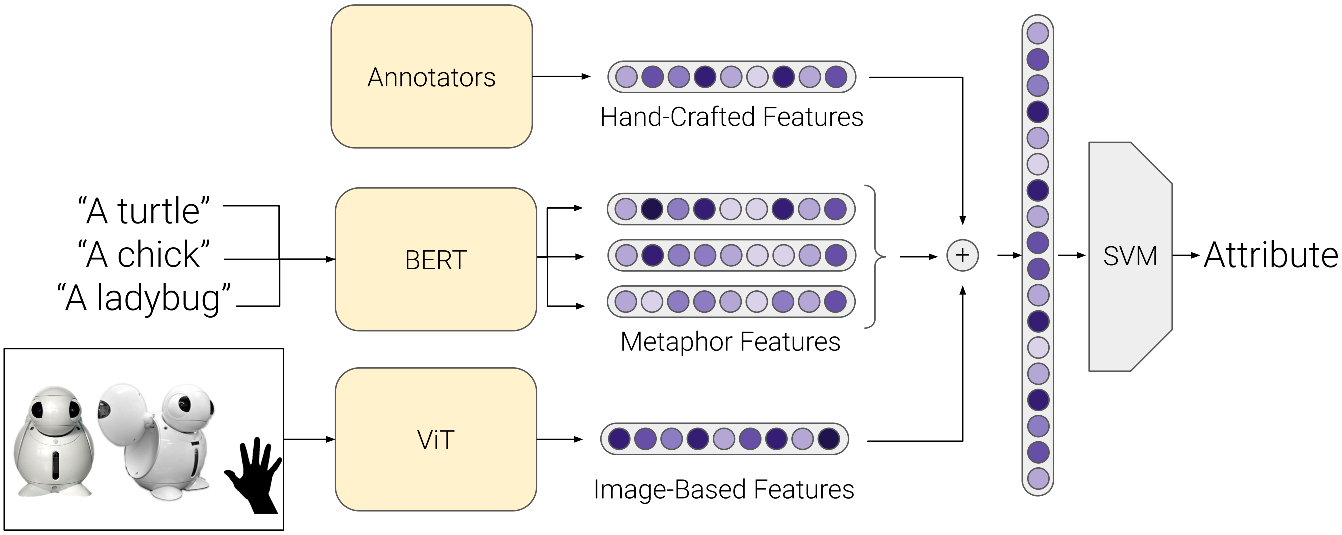

Fig. 1: Process for generating features from the MUFaSAA dataset.

A. MUFaSAA Dataset Description

The MUFaSAA Dataset is a collection of 165 socially interactive robots [10] that have been used by research labs or sold to consumers. Each robot has a uniform image representation that includes a front and side view with a height reference, a set of hand-coded design features (see [10] for feature descriptions), and a set of three design metaphors that participants used to describe the robot. Each robot also contains ratings of the three constructs from the validated Robotic Social Attributes scale [5]: Warmth, Competence, and Discomfort, and three constructs from the EmCorp-Scale [12]: Perception and Interpretation, Tactile Interaction, Nonverbal Communication. The values for these constructs are continuous values between -3 and 3 and represent the average rating of a 7-point Likert scales across approximately 30 participants for each robot.

B. Creating Features of Robot Embodiment

In this work, we generated three modes of features for each of the robots: Hand-crafted (HC) features, Metaphor (M) features, and Image-based (IM) features. The overall process we followed to generate features is shown in Figure 1. Metaphor features and Image-based features were deep features that came from large pre-trained models that were available from the transformers library.

1) Hand-crafted Features: Hand-crafted features were features of the robot's embodiment that previous research has found to be important for human robot interaction (e.g., height [23], waist-to-hip ratio [1, 2], presence of a mouth [14], etc.) as well as other features that participants used to describe the robot. These features were labeled by annotators that had access to images of the robot and other information from websites created by the robots' manufacturers. These values were all scaled to be between zero and one. For each robot, there were 59 HC features.

2) Metaphor Features: Metaphor features were created from the three metaphors that were most often used to describe each of the robots in the MUFaSAA dataset. These metaphors consisted of either a single noun, e.g., "a dog", "a kiosk", etc., or the name of a specific reference accompanied by context from where the reference is from, e.g., "Rosie the Robot from the Jetsons", "Eve from WALL-E", etc. These metaphors were converted to vectors using a BERT model pretrained on the Toronto BookCorpus [26] and English Wikipedia datasets and projected to a 512-dimensional space.

3) Image-based Features: Image-based features were created from the standardized images of the robots in the MUFaSAA dataset. The images were converted to vectors based on a pre-trained Vision Transformer (ViT) model that was pretrained on ImageNet-21k [9]. The pretrained model outputs were projected to a 512 dimension vector.

C. Regression Experiment

We formulated understanding user social and functional expectations as a series of regression problems. We used Support Vector Machines (SVM) to regress robot features onto each of the six constructs in the RoSAS and EmCorp scales. Experiments were conducted in the scikit-learn framework [22]. We selected SVMs because they are often used for datasets of this size [20], and empirically performed the best across all constructs compared to all other regression techniques implemented in scikit-learn. Ground-truth labels came from the user-reported values in the MUFaSAA dataset.

1) Model Details: The SVM regressor used the radial basis function as a kernel. The regularization hyperparameter, C, and the margin of tolerance hyperparameter, ϵ, were selected by performing a grid search over the discrete values [.001, .01, .1, 1, 10, 100] for both hyperparameters. These parameters were evaluated by their average mean squared error loss over all constructs and folds in a 20-fold cross-validation setting. The best-performing values were $C = 1.0$ and $\epsilon = 0.1$.

TABLE I: Regression Results.

Features Used	Warmth	Competence	Discomfort	Perception and Interpretation	Tactile Interaction	Nonverbal Communication
HC	0.145**	0.130*	0.306*	0.188*	0.381***	0.182***
M	0.209	0.163	0.401	0.262	1.390	0.387
IM	0.177	**0.119****	0.344	0.202*	0.445***	0.190***
HC + M	0.137**	0.134*	0.311*	0.184*	0.388***	0.182***
HC + IM	0.138*	0.122*	**0.303***	**0.182***	**0.337*****	0.174***
M + IM	0.183	0.122*	0.355	0.216	0.466***	0.187***
HC + M + IM	**0.135****	0.124*	0.307*	**0.182***	0.349***	**0.173*****
Baseline (Predict Mean)	0.208	0.176	0.398	0.278	1.42	0.452

All significance values calculated from a t-test with respect to the baseline over all folds. * denotes $p < .05$, ** denotes $p < .01$, *** denotes $p < .001$.

2) Evaluation: For our experiments, we calculated the average mean squared error (MSE) for each of the six constructs of interest in a 20-fold cross-validation setup. We compare our results to the baseline of predicting the average value for the constructs across all robots in the training set. We perform this evaluation with every combination of the modalities of describing embodiment that we outlined in Section III-B.

IV. Preliminary Results

The modeling results from the experiment are displayed in Table I. Our results show that ratings of social and functional constructs are associated with robot embodiments.

A. Single Mode Results

We found significant improvements over the baseline with only one mode of feature being used for the HC and IM features. There were no significant differences between the HC and IM features in regressing on the six constructs. This is of particular interest because it indicates that features used from frozen pre-trained networks can be as effective at predicting social and functional expectations of robots without the difficulties associated with annotation.

We did not observe any significant improvements using only metaphor information to predict robot expectations. This may be because the features that can be gained from language models do not contain information on the physical interactions that the metaphors have. Thus to gain more use from these metaphors, language models may need to ground their understanding of concepts in physical experience [3].

B. Multiple Mode Results

Nearly all combinations of modalities, except M+IM, showed significant improvements over the baseline. In general, the best performing methods involved combinations of multiple modes of features. This suggests that the different modes of features have complementary information that is useful in understanding users' mental models. However, these combinations didn't show significant improvements over single modes of features.

V. Future Work and Conclusion

This work introduces features of embodiment to understand how people form mental models of robots. While these results show that features of embodiment can be used to better understand social and functional expectations of robots, there are several ways that this work can be expanded. In particular, the text-based metaphor features were not as helpful for understanding expectations as the other features. Future work can explore alternate way to calculate these features that can include other information. The MUFaSAA dataset also contains information on frequency of metaphor responses and levels of abstraction that describe how closely the robot resembles each metaphor. This additional information may be leveraged to generate more informative features for understanding robot social expectations.

The methods and initial results presented here are a preliminary work that show the potential for features of embodiment to be useful for determining how robots are expected to behave. These results are also subject to other differences based on other external factors like social and cultural contexts as well as the actual interactions users have with the robots. While these results can be further refined, they show important relations between how robots are embodied and how they are expected to act, which is valuable to inform the physical and algorithmic design of future robots.

References

[1] Jasmin Bernotat, Friederike Eyssel, and Janik Sachse. Shape it–the influence of robot body shape on gender perception in robots. In *Social Robotics: 9th International Conference, ICSR 2017, Tsukuba, Japan, November 22-24, 2017, Proceedings 9*, pages 75–84. Springer, 2017.

[2] Jasmin Bernotat, Friederike Eyssel, and Janik Sachse. The (fe) male robot: how robot body shape impacts first impressions and trust towards robots. *International Journal of Social Robotics*, 13:477–489, 2021.

[3] Yonatan Bisk, Ari Holtzman, Jesse Thomason, Jacob Andreas, Yoshua Bengio, Joyce Chai, Mirella Lapata, Angeliki Lazaridou, Jonathan May, Aleksandr Nisnevich, et al. Experience grounds language. *arXiv preprint arXiv:2004.10151*, 2020.

[4] Julie Carpenter, Joan M Davis, Norah Erwin-Stewart, Tiffany R Lee, John D Bransford, and Nancy Vye. Gender representation and humanoid robots designed for domestic use. *International Journal of Social Robotics*, 1(3):261–265, 2009.

[5] Colleen M Carpinella, Alisa B Wyman, Michael A Perez, and Steven J Stroessner. The robotic social attributes scale (rosas) development and validation. In *Proceedings*

of the 2017 ACM/IEEE International Conference on human-robot interaction, pages 254–262, 2017.

[6] Elizabeth Cha, Anca D Dragan, and Siddhartha S Srinivasa. Perceived robot capability. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 541–548. IEEE, 2015.

[7] Fred D Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, pages 319–340, 1989.

[8] Eric Deng, Bilge Mutlu, Maja J Mataric, et al. Embodiment in socially interactive robots. *Foundations and Trends® in Robotics*, 7(4):251–356, 2019.

[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[10] Nathaniel Dennler, Changxiao Ruan, Jessica Hadiwijoyo, Brenna Chen, Stefanos Nikolaidis, and Maja Matarić. Design metaphors for understanding user expectations of socially interactive robot embodiments. *ACM Transactions on Human-Robot Interaction*, 12(2):1–41, 2023.

[11] Susan T Fiske, Amy JC Cuddy, and Peter Glick. Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences*, 11(2):77–83, 2007.

[12] Laura Hoffmann, Nikolai Bock, and Astrid M Rosenthal vd Pütten. The peculiarities of robot embodiment (emcorp-scale) development, validation and initial test of the embodiment and corporeality of artificial agents scale. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*, pages 370–378, 2018.

[13] Heekyoung Jung, Heather Wiltse, Mikael Wiberg, and Erik Stolterman. Metaphors, materialities, and affordances: Hybrid morphologies in the design of interactive artifacts. *Design Studies*, 53:24–46, 2017.

[14] Alisa Kalegina, Grace Schroeder, Aidan Allchin, Keara Berlin, and Maya Cakmak. *Characterizing the Design Space of Rendered Robot Faces*, page 96–104. Association for Computing Machinery, New York, NY, USA, 2018. ISBN 9781450349536. URL https://doi.org/10.1145/3171221.3171286.

[15] Pranav Khadpe, Ranjay Krishna, Li Fei-Fei, Jeffrey Hancock, and Michael Bernstein. Conceptual metaphors impact perceptions of human-ai collaboration. *arXiv preprint arXiv:2008.02311*, 2020.

[16] David E Kieras and Susan Bovair. The role of a mental model in learning to operate a device. *Cognitive science*, 8(3):255–273, 1984.

[17] Sara Kiesler and Jennifer Goetz. Mental models of robotic assistants. In *CHI'02 extended abstracts on Human Factors in Computing Systems*, pages 576–577, 2002.

[18] Jingoog Kim and Mary Lou Maher. Conceptual metaphors for designing smart environments: Device, robot, and friend. *Frontiers in Psychology*, 11:198, 2020.

[19] Minae Kwon, Malte F Jung, and Ross A Knepper. Human expectations of social robots. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 463–464. IEEE, 2016.

[20] Leena Mathur and Maja J Matarić. Introducing representations of facial affect in automated multimodal deception detection. In *Proceedings of the 2020 International Conference on Multimodal Interaction*, pages 305–314, 2020.

[21] Youngme Moon and Clifford Nass. How "real" are computer personalities? psychological responses to personality types in human-computer interaction. *Communication research*, 23(6):651–674, 1996.

[22] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

[23] Irene Rae, Leila Takayama, and Bilge Mutlu. The influence of height in robot-mediated communication. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 1–8. IEEE, 2013.

[24] Matthew Rueben, Jeffrey Klow, Madelyn Duer, Eric Zimmerman, Jennifer Piacentini, Madison Browning, Frank J Bernieri, Cindy M Grimm, and William D Smart. 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, 2021.

[25] Stephen Voida, Elizabeth D Mynatt, and W Keith Edwards. Re-framing the desktop interface around the activities of knowledge work. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, pages 211–220, 2008.

[26] Yukun Zhu, Ryan Kiros, Rich Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In *Proceedings of the IEEE international conference on computer vision*, pages 19–27, 2015.