Context^2: On the Importance of the Context of Context in Human Robot Interaction

Kate Candon
Computer Science Department
Yale University
New Haven, CT, USA
kate.candon@yale.edu

Marynel Vázquez
Computer Science Department
Yale University
New Haven, CT, USA
marynel.vazquez@yale.edu

***Abstract*—It is important for intelligent robots to be able to interpret human signals that provide context about how an interaction is going. We posit that including multiple facets of context, both situational and user-specific, in user models will improve a robot's understanding of the context of their interactions. This position is supported by results from an exploratory study where humans interacted with an agent in a video game. As part of this work, we built contextual perception models that reasoned about nonverbal human reactions to prosocial assistance from the autonomous agent. Interestingly, our results showed the importance of contextualizing model predictions based on multiple factors. Future work will further examine the importance of the inclusion of the context of context, or *context*^{2}, in perception models to make intelligent predictions about nonverbal reactions through richer utilization of our existing data. Additionally, we plan on extending our study to situated human-robot interactions.**

***Index Terms*—human-robot interaction; affective computing; context**

I. Introduction

As assistive robots become more prevalent in everyday life, it is important that they understand how they are being perceived by the people they interact with and can adapt their behavior accordingly. A human's internal state is not directly observable, but implicit feedback can provide hints about what a user may be thinking or feeling. Importantly, implicit feedback is provided "freely" during interactions, so it presents an opportunity to enrich a robot's observations at no additional burden to the user from an interaction perspective [1]. Implicit feedback encompasses a variety of user behaviors from which it is possible to infer feedback, ranging from facial expressions (e.g., [2]) to tone of voice (e.g., [3]).

Different definitions of *context* are used across, and even within, research areas. Examples include: the who-what-where-when of interactions [4]; global features extracted from visual images [5]; and information such as physical location, time of day, observable behaviors, and the social event taking place in an interaction [6]. Inspired by the definition of context awareness as the ability of an agent to perceive and react to aspects of their environment [7], we consider a broad definition of context that includes any element providing information about the environment in which an agent is situated. For example, we consider nonverbal human reactions, demographics and personality traits of users, and activity statistics to all be contextual factors relevant to human-robot interaction.

This paper posits that considering the context of context, or *context*^{2}, of human-robot interactions can help create more adaptive social robots. For example, consider a robot that assists a user who is preparing a meal. If the robot perceives a smile after it has just handed the user a spoon, it could signal that the user is happy with the assistance the robot is providing. On the other hand, if the robot has not recently provided any help but is just watching the user, a similar smile could signal the user feels awkward about being watched. In both scenarios, the perceived smile provides contextual information to the robot, but considering additional context about the interaction could allow the robot to better interpret the smile, and thus better reason about the internal state of the user.

More generally, this paper argues in favor of regarding context as a complex and multi-faceted construct. A lot of potential information could be lost by focusing on individual aspects. Rather, in the Human-Robot Interaction community, we should consider context as a collection of different factors. Each factor has the potential to help us better understand both other contextual elements and the situation in which the robot operates.

II. Related Work

A. Helpfulness and Prosocial Help by Autonomous Agents

Researchers have studied what makes an autonomous agent a good collaborator. With respect to assistance, the timing of helping actions is important [8]. Other work has highlighted the nuance of understanding how assistance is received. For example, prior work has found that actions need not only be objectively useful but also perceived to be useful [9]. Other work has found that helping actions taken by an agent may not necessarily be perceived as assistance [10]. Additional work has highlighted the importance of reasoning about multi-dimensional aspects of teamwork, such as displaying effort, for agents to be considered helpful [11].

Understanding whether an autonomous agent is helpful becomes even more difficult when humans are not primed for a specific type of cooperation. This type of assistance is known as prosocial assistance: when one agent takes an

action that benefits others despite some personal cost [12]. The question of how to create social, computational agents that are capable of rendering prosocial actions has inspired significant work within artificial intelligence [13], [14]. Prior work has had success with leveraging task-specific variables to encourage prosocial behavior, but has struggled to understand more individual indicators [15]. We suspect that one avenue to gain insight into these individual indicators is by analyzing nonverbal human reactions.

B. Implicit Feedback

There has been significant effort within social signal processing to computationally model and understand nonverbal human behavior [4]. For example, researchers have used computational models of nonverbal human behavior to adapt the behavior of in-home devices [16]. Recently, there has been evidence of successfully adapting the behavior of agents via reinforcement learning with this kind of implicit feedback [2], though participants were only observing the agent, rather than interacting with it. Similarly, the inclusion of social cues via facial expressions has been shown to improve a generative deep learning model [17].

There has been recent discourse about the limitations of interpreting facial expressions [18]. For example, emotions are represented differently across cultures and context [19]. We believe that including context, in the sense of activity statistics as well as demographics and personality, will improve the quality of predictions made about nonverbal reactions in human-robot interactions.

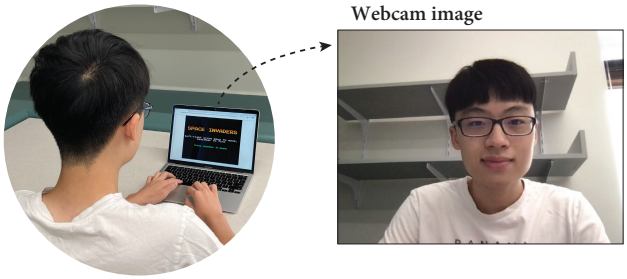
C. Context in Human Robot Interaction

Context has been shown to be an integral part of understanding human-robot interactions [20]. For example, the definition of trust in human-robot interactions may hinge on the context of the interaction [21]. Context can also improve a robot's ability to understand user attention [22] or solve perception tasks [6]. Related to our position, interactional context has been shown to improve classification of human gestures compared to kinematics alone [23]. Additionally, researchers have illustrated the success of leveraging additional contextual information from images in emotion recognition [24]. We see an opportunity to enrich the notion of context to include additional contextual information, such as personality traits of a user and activity statistics of an interaction, in order to better understand nonverbal human reactions to an autonomous agent.

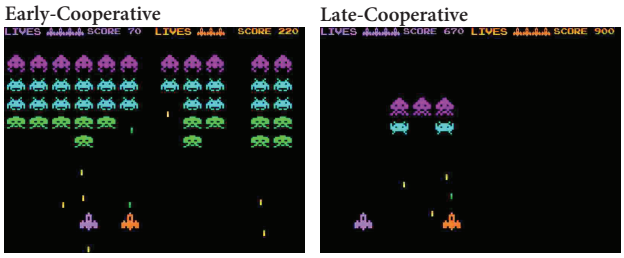
III. Exploratory Study

Our position is motivated by previous, under-review work in which we conducted an exploratory online study with 194 participants to explore how people perceive and react to prosocial help from an autonomous agent in a multiplayer version of the Space Invaders game. This section gives a brief overview of our previous work to motivate our position.

In our Space Invaders game, each participant controlled a spaceship that spawned on the left side of the game screen. The autonomous agent, or "co-player", controlled the spaceship that spawned on the right side of the game screen. Both players received points for the enemies destroyed on the side on which they were spawned. Our experimental setup and our version of the Space Invaders game can be seen in Figure 1.



(a) Experimental setup: The participants' faces were recorded while playing the Space Invaders game in our online study.



(b) The participants experienced two types of helping behaviors by the co-player (orange ship) in the study.

Fig. 1. We conducted a study to investigate how participants perceived a prosocial agent in a multi-player Space Invaders game. We controlled for the agent's helping behavior (early-cooperative vs. late-cooperative).

In our study, the participant experienced two different behaviors from the co-player. In one game, the co-player exhibited an early-cooperative behavior, in which the agent went over to the participant's side of the game screen and helped destroy enemies on two separate occasions before it had finished destroying its own enemies on the right side of the game screen. In the other game, the co-player exhibited a late-cooperative behavior. This agent only went over to the left side of the game screen to help destroy enemies after it had destroyed all of its own enemies on the right side of the game screen. Figure 1(b) illustrates examples of the two different helping behaviors. By destroying enemies on the left side of the game screen, the co-player was helping the participant achieve their objective of scoring points for enemies destroyed on their side. However, we found that helping to destroy left enemies did not necessarily translate to helpfulness ratings nor the participant reporting that they liked or preferred a particular behavior.

Our preliminary results suggested there may not be universal truths when it comes to understanding how an agent's help will be received, nor what actions we can assume humans will interpret as helpful. Rather, it is important to understand the

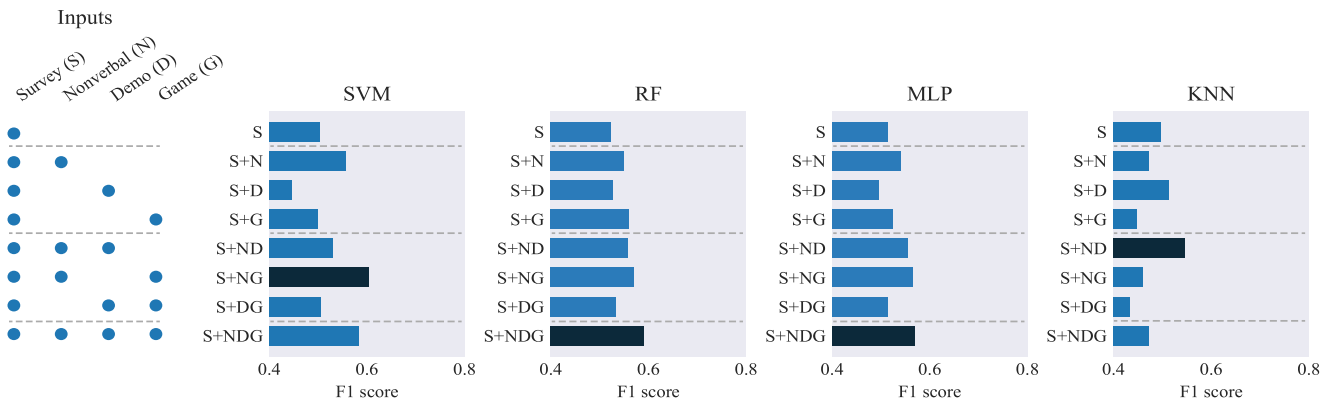


Fig. 2. Best F_1-Scores for each combination of information inputs. F_1-Score was calculated over confusion matrix derived from individual predictions of 194 folds of LOOCV. The dots on far-left indicate which information was considered in the model. Machine learning algorithms (Support Vector Machine (SVM), Random Forest (RF), Multi-Layer Perceptron (MLP), K-Nearest Neighbors (KNN)) are ordered left-to-right in decreasing order of highest F_1-Score. The darkest bar highlights the highest F_1-Score for each algorithm. Dotted lines separate input combinations into: no additional interaction context, one type of additional interaction context, two types of additional interaction context, and all three types of additional interaction context.

individual and to adapt to what is influencing their perception of the interaction and the type of behavior that they prefer.

Because we anticipated there may be nuance in how prosocial actions would be perceived, we were also interested in analyzing implicit feedback in the form of nonverbal reactions. To this end, we recorded the participants' upper bodies via their webcam while they played the two games of Space Invaders. In particular, we explored if implicit feedback and other context could be leveraged to understand user preferences for an agent's helping behaviors without explicitly biasing humans towards expressivity. To investigate this, we considered the participants' stated preferences between the two helping behaviors as a multi-class classification problem with targets in the set: $\{$*Early, Late, No Preference*$\}$.

We explored incorporating additional contextual information into classification models. The base input data consisted of features derived from post-game survey responses. We added additional context to the survey data (S) via combinations of three types of contextual information: 1) nonverbal reaction data (N) via various summary statistics of facial action unit and body motion features extracted via OpenFace 2.0 [25], 2) participant-provided demographic data (D), and 3) game context (G) via game logs. We suspected possible information leakage from the survey responses and hypothesized that additional contextual information would improve our ability to predict user preferences over agent behavior.

Our results support the idea that considering multiple facets of context allows an agent to better reason about the internal state of a user. In Figure 2, the highest F_1-Score for each algorithm included nonverbal reaction data and at least one other kind of additional context. Notably, including all three kinds of additional contextual information did not always result in the highest F_1-Score (e.g., SVM and KNN), so one cannot take for granted that including more features will result in higher performance. Rather, it is important to further explore how to best incorporate the context of context when reasoning about internal human states.

IV. Future Directions

Our preliminary results align with recent research in social psychology that contests the assumption that emotions are recognized and communicated universally with particular facial expressions and argues for the importance of considering context when analyzing human behavior (e.g., [18], [19]). We think it is imperative to continue to study the importance of context for interpreting nonverbal human reactions to autonomous, interactive behavior. In the future, we plan to explore richer utilization of the data from our exploratory study, and will extend our analysis to human-robot interactions.

One future direction is to enable robots to use implicit feedback with context as a continuous reward to understand how their behaviors are being perceived by a user. We can describe an interaction between a human and a robot as a sequence of states (s) and actions taken by the human (a_H) and robot (a_R): $\xi = \langle\{s^0, a_H^0, a_R^0\}, ..., \{s^T, a_H^T, a_R^T\}\rangle$. The state, $s = [s_E, s_R, s_H]^T$, includes information about the environment (s_E), the robot (s_R), and the internal state of the human (s_H). We will explore using machine learning methods to learn a reward function r that maps state-action tuples (s, a_R, a_H) to a value in $[0, 1]$ indicating how positively the human perceived the robot action a_R in the context of the state s when the human took action a_H. With fully observable data, it is possible to learn the reward function r to build an adaptable robot [26]. However, the internal state of the human (s_H) is not fully observable, so we will study how to learn to assign meaning to observable implicit signals, which could enable a robot to make predictions about the internal human state (s_H).

Another future direction is to extend our perception models to interpret nonverbal human reactions in context during

situated human-robot interactions. We plan to conduct an in-person study similar to our preliminary online exploratory study, which will introduce additional complexities. For example, participants may be distracted by the robot, compared to playing against a hypothetical co-player online. A situated environment could also introduce noise to our data. One example is that smiling and laughing are common reactions to unexpected robot behavior [27]. However, an embodied robot introduces the possibility that participants would be more expressive, providing additional data to analyze. The robot could also take error recovery strategies during the interaction that could be considered as additional context.

Our future work aims to improve situated human-robot interactions by enabling robots to understand how their actions are being perceived in real-time and allowing them to adapt their behavior to the preferences of individual users. Additionally, this future work could help us to better understand the actions that tend to prompt informative nonverbal human reactions and how to incorporate those actions into a robot's behavior earlier in interactions.

V. Conclusion

We posit that including multiple levels of contextual information into decision-making models will improve a robot's ability to understand the context of its interactions with humans. Researchers have thought of nonverbal human reactions as contextual factors to reason about the capabilities of robots (e.g., [2]). In our Space Invaders game, we similarly considered human nonverbal reactions as contextual clues about what kind of helping behavior participants preferred from an autonomous agent. We found that considering additional contextual factors, such as information about the game state and participant demographics and personality traits, improved our ability to reason about nonverbal human feedback to predict participant preferences across helping behaviors. Thus, we encourage the Human-Robot Interaction community to consider a robust collection of features as context. Considering the context of context, or *context*^{2}, should enable intelligent robots to make fuller use of implicit human feedback gathered during interactions.

Acknowledgment

Thank you to the National Science Foundation (IIS-2106690) for partially supporting this work.

References

[1] J. Lin, Z. Ma, R. Gomez, K. Nakamura, B. He, and G. Li, "A review on interactive reinforcement learning from human social feedback," *IEEE Access*, vol. 8, pp. 120 757–120 765, 2020.

[2] Y. Cui, Q. Zhang, A. Allievi, P. Stone, S. Niekum, and W. B. Knox, "The empathic framework for task learning from implicit human feedback," in *Proceedings of the 4th Conference on Robot Learning (CoRL 2020)*, Cambridge MA, USA, November 2020.

[3] E. S. Kim and B. Scassellati, "Learning to refine behavior using prosodic feedback," in *2007 IEEE 6th International Conference on Development and Learning*. IEEE, 2007, pp. 205–210.

[4] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," *Image and vision computing*, vol. 27, no. 12, pp. 1743–1759, 2009.

[5] R. Kosti, J. M. Alvarez, A. Recasens, and A. Lapedriza, "Emotion recognition in context," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1667–1675.

[6] A. Nigam and L. D. Riek, "Social context perception for mobile robots," in *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2015, pp. 3621–3627.

[7] B. N. Schilit and M. M. Theimer, "Disseminating active map information to mobile hosts," *IEEE network*, vol. 8, no. 5, pp. 22–32, 1994.

[8] L. C. Jensen, K. Fischer, S.-D. Suvei, and L. Bodenhagen, "Timing of multimodal robot behaviors during human-robot collaboration," in *RO-MAN 2017*. IEEE, 2017, pp. 1061–1066.

[9] P. Rudman and M. Zajicek, "Autonomous agent as helper-helpful or annoying?" in *2006 IEEE/WIC/ACM International Conference on Intelligent Agent Technology*. IEEE, 2006, pp. 170–176.

[10] J. Kim, Y. J. Sah, and H. Song, "Agreeableness of a virtual agent: Effects of reciprocity and need for help," in *2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM)*. IEEE, 2021, pp. 1–6.

[11] M. L. Chang and A. Thomaz, "Valuable robotic teammates: Algorithms that reason about the multiple dimensions of human-robot teamwork," in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 2021, pp. 580–582.

[12] J. F. Dovidio, J. A. Piliavin, D. A. Schroeder, and L. A. Penner, *The social psychology of prosocial behavior*. Psychology Press, 2017.

[13] C. Castelfranchi, "Modelling social action for ai agents," *Artificial intelligence*, vol. 103, no. 1-2, pp. 157–182, 1998.

[14] A. Paiva, F. Santos, and F. Santos, "Engineering pro-sociality with autonomous agents," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[15] R. Oliveira, P. Arriaga, F. P. Santos, S. Mascarenhas, and A. Paiva, "Towards prosocial design: A scoping review of the use of robots and virtual agents to trigger prosocial behaviour," *Computers in Human Behavior*, vol. 114, p. 106547, 2021.

[16] Y. Yan, C. Yu, W. Zheng, R. Tang, X. Xu, and Y. Shi, "Frownonerror: Interrupting responses from smart speakers by facial expressions," in *Proceedings of the 2020 CHI Conference*, 2020, pp. 1–14.

[17] N. Jaques, J. McCleary, J. Engel, D. Ha, F. Bertsch, D. Eck, and R. Picard, "Learning via social awareness: Improving a deep generative sketching model with facial feedback," in *Workshop on Artificial Intelligence in Affective Computing*. PMLR, 2020, pp. 1–9.

[18] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak, "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements," *Psychological science in the public interest*, vol. 20, no. 1, pp. 1–68, 2019.

[19] R. E. Jack, O. G. Garrod, H. Yu, R. Caldara, and P. G. Schyns, "Facial expressions of emotion are not culturally universal," *Proceedings of the National Academy of Sciences*, vol. 109, no. 19, pp. 7241–7244, 2012.

[20] L. Onnasch and E. Roesler, "A taxonomy to structure and analyze human–robot interaction," *International Journal of Social Robotics*, pp. 1–17, 2020.

[21] D. Cameron, J. Aitken, E. Collins, L. Boorman, A. Chua, S. Fernando, O. McAree, U. Martinez Hernandez, and J. Law, "Framing factors: The importance of context and the individual in understanding trust in human-robot interaction," 2015.

[22] H. Salam and M. Chetouani, "A multi-level context-based modeling of engagement in human-robot interaction," in *2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG)*, vol. 3. IEEE, 2015, pp. 1–6.

[23] C. L. Nehaniv, K. Dautenhahn, J. Kubacki, M. Haegele, C. Parlitz, and R. Alami, "A methodological approach relating the classification of gesture to identification of human intent in the context of human-robot interaction," in *RO-MAN 2005*. IEEE, 2005, pp. 371–377.

[24] J. Lee, S. Kim, S. Kim, J. Park, and K. Sohn, "Context-aware emotion recognition networks," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 10 143–10 152.

[25] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*. IEEE, 2018, pp. 59–66.

[26] D. Sadigh, A. D. Dragan, S. Sastry, and S. A. Seshia, "Active preference-based learning of reward functions," 2017.

[27] S. Honig and T. Oron-Gilad, "Understanding and resolving failures in human-robot interaction: Literature review and model development," *Frontiers in psychology*, vol. 9, p. 861, 2018.