Social Errors in Human-Robot Interaction

Leimin Tian
Faculty of Information Technology
Monash University
Melbourne, Australia
Leimin.Tian@monash.edu

Sharon Oviatt
Faculty of Information Technology
Monash University
Melbourne, Australia
Sharon.Oviatt@smonash.edu

***Abstract*—In recent years, robotic applications have entered various aspects of our lives, including health care and educational services. These domains require long-term Human-Robot Interaction (HRI), in which trust and mutual adaptation is established and maintained through a positive social relationship between the robot and the user. Such social bond relies on the perceived competence of the robot on the social-emotional dimension. However, because of technical limitations and user heterogeneity, current HRI is far from error-free, especially when the system leaves controlled lab environments and is applied to natural, everyday environments.**

To better understand the impact of errors in HRI and effective strategies to handle such impact, we propose to classify errors in HRI into two categories: performance errors which degrade the user's perception of the robot's intelligence and capability in achieving a task, and social errors which degrade the user's perception of the robot's social skills and their relationship with the robot. We focus on social errors in HRI and propose an operational definition to it. We argue that by addressing social errors in HRI, we can leverage them into opportunities for improving the socio-affective competence of HRI systems. Our work will contribute to identifying effective error handling strategies which lead to more personalized, adaptive, and socially acceptable interaction experiences in long-term HRI.

***Index Terms*—Affective computing, social robotics, human-robot interaction, social norms, socio-affective competence, personlization, long-term HRI**

I. Introduction

With recent technical advances, we have witnessed significant growth in the application of Artificial Intelligence (AI) in various domains of our society, such as educational or medical applications (see [1], [2] for reviews). Because of the importance of emotions in human cognition and communication [3], it is inevitable for AI researchers to take emotions into account, which led to the establishment of the research field Affective Computing [4]. The term *Affective* here refers to aspects of cognition relating to, resulting from, or influenced by human emotions, and Affective Computing aims at developing emotion-aware technologies.

The majority of current Affective Computing studies have been focused on automatic emotion recognition using various Machine Learning approaches. However, there remains considerable gap between performance of automatic emotion recognition models and human performance. Moreover, when applying such emotion recognition models trained on lab-collected databases to a more natural and spontaneous scenario, performance of the emotion recognition model potentially worsens (see [5], [6] for reviews). Beyond emotion recognition, identifying and expressing appropriate reactions to the recognized emotions is also vital for realizing emotion-aware interactions, and it remains an unsolved topic in Affective Computing [7].

Even with the limited performance of current emotion recognition and interaction functions, emotion-aware Human-Robot Interaction (HRI) has been found to foster and enhance human-robot relationship and leads to more personalized and adaptive interaction experiences in various studies [8]–[10]. Moreover, emotional responsiveness and interpersonal warmth are key factors influencing people's perception of the robot [11] and the outcome of HRI applications. For example, in senior care and healthy aging domain, HRI designs which incorporate social-emotional aspects in the interaction have been shown to improve the health outcomes by encouraging positive moods and reducing loneliness felt by the users [12]. In collaborative HRI and human-robot teamwork, the socio-affective competence of the robot can largely influence people's trust towards the robot, which is critical to human's decision-making and willingness to cooperate, especially in uncertain or risky situations [13]. In long-term and situated HRI scenarios where the effects of novelty reduce over time [14], such as home assistant robots [15], perceived socio-affective competence of the robot is key to establishing a social relationship between the user and the robot [16], [17].

HRI has been shown to activate emotional reactions and psychological mechanisms in the user comparable to human-human interactions [18]. Thus, there have been growing studies in utilizing HRI for mental health care. For example, social robots have been used to help children with autism spectrum disorders to understand social cues and practice interpersonal interactions, which benefits their social inclusion and quality of life (see [19], [20] for reviews). Similarly, the assistance of HRI in elderly care has also provided valuable support to the basic social-emotional needs of the elderly citizens (see [12], [21] for reviews). Because of the known benefits of social relationships, such as reducing morbidity [22] and mortality [23], HRI systems which can establish and maintain a social relationship with the user or encourage the user to be involved in social relationships with other people have great potentials in health and well-being applications.

With current social-emotional interaction functions of HRI far behind human-level performance, errors are inevitable

during the interaction. Such errors may have significant impact on the user's perception of the robot and the HRI. However, error handling is yet to be fully understood in current HRI studies, especially regarding errors on the social-emotional dimension of HRI. Previous research on errors in HRI have been focused on performance errors in the functional components of the robot, such as navigation [24]. However, violation of social norms was shown to cause changes in behavioral and neural reactions in human studies [25]. Moreover, previous Psychology studies have found that compliance and adaptation to mutually agreed social norms is essential to social bonding and establishing relationships [26]. These findings suggest that social errors made by the robot can have significant influence on the user's perception of the robot and the interaction. Therefore, understanding the influence of social errors will be a key to advance current HRI research, especially in long-term interaction scenarios where a positive human-robot relationship is desired, such as elderly care. This motivates us to bridge the gap in current HRI research by formally defining social errors in HRI, and providing a systematic analysis on key attributes and impact of such errors.

Moreover, errors are opportunities for improvement. For example, a recent study on learning-based approach to hand-eye coordination for robotic grasping has found that more reliable and effective grasping can be learned through correcting mistakes [27]. Therefore, by addressing social errors in HRI, it is reasonable to believe that we can identify effective error handling strategies which lead to more personalized, adaptive, and socially acceptable HRI. This is crucial for achieving objective outcomes of HRI applications, such as better study outcomes in educational applications, more humane healthcare, and more effective human-robot collaborative problem solving. Our study will contribute to current understandings of the social-emotional aspects of HRI and serve as a foundation for advancing socio-affective competence of current HRI systems.

II. Social errors and performance errors in HRI

We classify errors in HRI into two types with the following operational definitions:

- **Performance errors** are errors which degrade the user's perception of the robot's intelligence and competence in achieving a task, such as failure to register a spoken command given by the user in a noisy environment; and
- **Social errors** are errors which violate social norms and degrade the user's perception of the robot's socio-affective competence and their relationship with the robot, such as interrupting the user at an inappropriate time during a conversation.

A social error may be caused by technical failures, such as delayed dialogue responses, or by imperfect design of the social-emotional interaction functions of the HRI system, such as a rule-based emotion interaction model not addressing individual variances. An error scenario can be a mix of performance error and social error. For example, a robot not able to register a spoken command repeatedly may irritate the user and result in abortion of the interaction. In this study, we focus on understanding the *impact* of these errors, rather than their causes. Note that social errors are context-dependent. The same HRI scenario can be perceived differently and cause entirely different impact due to individual variances, cultural differences, mental health status, and various contextual factors. Our discussion of social errors in HRI will be situated under these environmental and contextual factors.

The main research question we address here is how to systematically analyze social errors in HRI. Our hypothesis is that we can develop a taxonomy of social errors applicable to HRI based on psychology and social cognition theories of human interpersonal interactions, and this taxonomy can guide systematic analysis of the impact of social errors on the perceived socio-affective competence of the robot and the social relationship between the robot and its human user.

Various definitions of socio-affective competence and social relationship exist in the literature of psychology, psychiatry, social science, and economy. In this study, we adopt the definition of socio-affective competence as the ability to successfully conduct social interactions, which depends on the awareness and identification of social-emotional cues, the ability to process such cues, and the ability to decide on and express a normative response to these cues [28], [29]. We adopt the definition of social relationship as a connection between a person and another entity, which represents the person's perception of the availability or adequacy of resources provided by the other entity, and results in interdependency of their social behaviors [30], [31].

III. Existing taxonomies of errors in HRI

Most previous research on human's perception of errors in HRI did not distinguish between performance errors and social errors. For example, a recent study listed a set of scenarios in which a domestic service robot has erratic behaviors, and collected ratings on people's perception of severity of these errors [32]. These scenarios are a mix of performance errors and social errors by our definition, and there were both types of errors being perceived as severe. To better understand errors in HRI, several taxonomies of errors have been proposed in previous studies. These taxonomies described errors using attributes include functional severity, social severity, relevance, frequency, condition, and symptoms [33]:

- Laprie [34] classified errors into two types by severity: benign errors (consequences of errors are comparable to the benefits of the service), and catastrophic errors (consequences of errors have a higher cost by one or more orders of magnitude than the benefits).
- Ross et al. [35] classified errors into four types by recoverability: anticipated errors (the robot can backtrack through the original plan to achieve the original goal through an alternate action sequence), exceptional errors (the original plan cannot cope with the failure, but with re-planning the original goal can still be achieved), unrecoverable errors (the original goal cannot be achieved

either through backtracking or by re-planning), and socially recoverable errors (the robot can continue on with the original plan with appropriate assistances from its environment).
- Carlson and Murphy [36] classified errors into physical errors and human errors first, then further classified physical errors by severity and recoverability, and classified human errors as design errors and interaction errors.
- Steinbauer [37] classified errors into four types: interaction errors, algorithm errors, software errors, and hardware errors.
- Brooks [38] classified errors into two types: communication errors, and processing errors.

The most closely related HRI error taxonomy is the work of Giuliani et al. [39]. They analyzed user behaviors during four error situations in multiple HRI studies, namely long dialogue pauses, repetitions in the dialogue, misunderstandings, and complete abruption of the interaction. Human annotators classified the errors into technical failures and violations of social norms. They defined social norm violation as a deviation from the social script or the usage of the wrong social signals. In particular, they studied violations of social norms as the robot executing interaction steps at the wrong time, or showing unusual social signals. They collected manual annotations of five types of user actions in the HRI sessions when an error situation occurs, as listed below:

- Spoken sentence types: task-related sentences, questions, statements, laugh, and correction.
- Head movements: the participant looks at the robot, at a group member, into a direction, or at the experimenter; nodding, shaking, or tilting the head.
- Facial expressions: smile, raise eyebrows, grimace.
- Body movements: leaning towards or away from the robot, moving towards or away from the robot, and change of posture.
- Hand gestures: self-touching, manipulating an object, and pointing.

In their analyses, Giuliani et al. compared the average number of occurrences of each action in an interaction session labeled as containing social norm violation or technical failure. They further compared different interaction settings, namely the experimenter being visible or not, and single user vs. group interactions. They found different user behavioral patterns when different types of error occurred. This supports our claim to distinguish social errors and performance errors in HRI. However, Giuliani et al. focused on visual-based behavioral analysis. To better understand the impact of errors in HRI, we are motivated to conduct more detailed user analyses. For example, collecting self-evaluations from the user regarding their perception of the robot when errors occur, or conducting quantitative studies to measure the impact of social errors on the user's trust towards the robot. Moreover, only limited, dialogue-oriented error scenarios were studied by Giuliani et al. We plan to extend our study of errors in HRI to beyond dialogue-based interaction scenarios.

IV. Future directions

Being a work-in-progress, in the current stage, we are reviewing literatures to design the taxonomy of social errors in HRI. In particular, we are reviewing psychological theories of emotions, empathy, socio-affective competence, and social norms. We are also reviewing current HRI research relevant to our study, including human perception of HRI errors, and adaptation and personalization in long-term HRI. Based on our reviews, we will propose a taxonomy of social errors in HRI, which reflects the major attributes of socio-affective competence in interpersonal interactions. Following the taxonomy, we will design HRI scenarios addressing each attribute of social errors.

In the experimental stage, we will first conduct crowdsourced surveys to analyze human perception of the designed HRI scenarios. This allows us to gather initial assessments on human perception of social errors in HRI. Moreover, it provides evidences for us to refine our HRI scenario designs, and identify any potential ethical concerns.

After these simulation studies, we will implement the HRI sessions using a Pepper robot [40] and conduct HRI experiments. For complex HRI scenarios or error-free conditions, the Wizard-of-Oz approach will be used where the robot is remotely controlled by a hidden human operator. The HRI experiments will include both short-term, single-session interactions, and long-term, longitudinal interactions. For example, placing the robot at a reception desk where it carries out basic receptionist duties serving both first-time visitors and regular visitors, such as answering questions or delivering mails. The interaction sessions will be designed to examine the impact of different types of social errors addressing each attribute of human perception of socio-affective competence, such as accuracy of the emotion recognition function of the robot.

In short-term HRI experiments, the participants will answer questionnaires before and after the experiments reflecting their perception of the robot and the HRI session. We will also annotate quantitative measurements, such as task success rate or engagement level of the participants during the HRI session. The interaction sessions will be recorded for detailed analysis of the behaviors of participants, such as changes in their gaze or speech when social errors occur. In long-term HRI experiments, we will examine the efficacy of failure recovery strategies and their impacts on the personalization of HRI and the perceived social relationship between the participants and the robot. For example, requesting user input when the robot is unable to recognize the meaning of a facial expression of the person, and applying the information learned in future interaction sessions with this person. We will examine how such a personalized and adaptive HRI system influences its user's perceptions and behaviors with both quantitative and qualitative analysis. For example, having the user and the robot collaborating in the same task, such as map navigation, at different stages of the longitudinal HRI.

Our systematic analysis on the impact of social errors in HRI will serve as the foundation for developing personal,

adaptive, and socially-acceptable long-term HRI applications. More importantly, we hope our work will inspire and facilitate discussion in the HRI research community regarding, but not limited, to the following topics:

- What are the social-emotional impacts of errors in short-term and longitudinal HRI?
- Instead of viewing errors as destructive events that should be eliminated, do they possess any information that can be utilized as well?
- Can we achieve personalized and adaptive HRI by addressing social errors in HRI?

REFERENCES

[1] D. Y. Y. Sim and C. K. Loo, "Extensive assessment and evaluation methodologies on assistive social robots for modelling human-robot interaction - a review," *Information Sciences*, vol. 301, pp. 305–344, 2015.

[2] T. B. Sheridan, "Human–robot interaction: Status and challenges," *Human factors*, vol. 58, no. 4, pp. 525–532, 2016.

[3] L. Pessoa, "Emotion and the interactive brain: Insights from comparative neuroanatomy and complex systems," *Emotion Review*, vol. 10, no. 3, pp. 204–216, 2018.

[4] R. W. Picard, *Affective computing*. MIT press, 2000.

[5] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," *Information Fusion*, vol. 37, pp. 98–125, 2017.

[6] M. H. Pestana, W.-C. Wang, and L. Moutinho, "The knowledge domain of affective computing: A scientometric review," in *Innovative Research Methodologies in Management*. Springer, 2018, pp. 83–101.

[7] A. Arora, "Action model learning for socio-communicative human robot interaction," Ph.D. dissertation, Université Grenoble Alpes, 2018.

[8] A. Arora, H. Fiorino, D. Pellier, and S. Pesty, "A review on learning planning action models for socio-communicative HRI," in *Workshop on Affect, Compagnon Artificiel and Interaction*, 2016.

[9] S. Thunberg, S. Thellman, and T. Ziemke, "Don't judge a book by its cover: A study of the social acceptance of NAO vs. Pepper," in *Proceedings of the 5th International Conference on Human Agent Interaction*. ACM, 2017, pp. 443–446.

[10] L. Paletta, M. Fellner, S. Schüssler, J. Zuschnegg, J. Steiner, A. Lerch, L. Lammer, and D. Prodromou, "AMIGO: Towards social robot based motivation for playful multimodal intervention in dementia," in *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*. ACM, 2018, pp. 421–427.

[11] N. Haslam, "Dehumanization: An integrative review," *Personality and social psychology review*, vol. 10, no. 3, pp. 252–264, 2006.

[12] S. Góngora Alonso, S. Hamrioui, I. de la Torre Díez, E. Motta Cruz, M. López-Coronado, and M. Franco, "Social robots for people with aging and dementia: A systematic review of literature," *Telemedicine and e-Health*, 2018.

[13] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. Chen, E. J. De Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Human Factors*, vol. 53, no. 5, pp. 517–527, 2011.

[14] C. A. Kubota and R. G. Olstad, "Effects of novelty-reducing preparation on exploratory behavior and cognitive learning in a science museum setting," *Journal of research in Science Teaching*, vol. 28, no. 3, pp. 225–234, 1991.

[15] I. Leite, C. Martinho, and A. Paiva, "Social robots for long-term interaction: A survey," *International Journal of Social Robotics*, vol. 5, no. 2, pp. 291–308, 2013.

[16] T. W. Bickmore and R. W. Picard, "Establishing and maintaining long-term human-computer relationships," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 12, no. 2, pp. 293–327, 2005.

[17] D. Feil-Seifer and M. J. Matarić, "Toward socially assistive robotics for augmenting interventions for children with autism spectrum disorders," in *Experimental robotics*. Springer, 2009, pp. 201–210.

[18] M. F. Jung, N. Martelaro, and P. J. Hinds, "Using robots to moderate team conflict: the case of repairing violations," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2015, pp. 229–236.

[19] P. Pennisi, A. Tonacci, G. Tartarisco, L. Billeci, L. Ruta, S. Gangemi, and G. Pioggia, "Autism and social robotics: A systematic review," *Autism Research*, vol. 9, no. 2, pp. 165–183, 2016.

[20] S. Golestan, P. Soleiman, and H. Moradi, "A comprehensive review of technologies used for screening, assessment, and rehabilitation of autism spectrum disorder," *arXiv preprint arXiv:1807.10986*, 2018.

[21] A. Korchut, S. Szklener, C. Abdelnour, N. Tantinya, J. Hernández-Farigola, J. C. Ribes, U. Skrobas, K. Grabowska-Aleksandrowicz, D. Szczęśniak-Stańczyk, and K. Rejdak, "Challenges for service robots—requirements of elderly adults with cognitive impairments," *Frontiers in neurology*, vol. 8, p. 228, 2017.

[22] B. S. Wallston, S. W. Alagna, B. M. DeVellis, and R. F. DeVellis, "Social support and physical health," *Health psychology*, vol. 2, no. 4, p. 367, 1983.

[23] J. S. House, C. Robbins, and H. L. Metzner, "The association of social relationships and activities with mortality: Prospective evidence from the Tecumseh community health study," *American journal of epidemiology*, vol. 116, no. 1, pp. 123–140, 1982.

[24] M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn, "Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2015, pp. 141–148.

[25] M. Spitzer, U. Fischbacher, B. Herrnberger, G. Grön, and E. Fehr, "The neural signature of social norm compliance," *Neuron*, vol. 56, no. 1, pp. 185–196, 2007.

[26] P. A. Thoits, "Emotion norms, emotion work, and social order," in *Feelings and emotions: The Amsterdam symposium*. Cambridge University Press New York, NY, 2004, pp. 359–378.

[27] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," *The International Journal of Robotics Research*, vol. 37, no. 4-5, pp. 421–436, 2018.

[28] A. G. Halberstadt, S. A. Denham, and J. C. Dunsmore, "Affective social competence," *Social development*, vol. 10, no. 1, pp. 79–119, 2001.

[29] N. Eisenberg, "The core and correlates of affective social competence," *Social development*, vol. 10, no. 1, pp. 120–124, 2001.

[30] H. B. Bosworth and K. W. Schaie, "The relationship of social environment, social networks, and health outcomes in the seattle longitudinal study: Two analytical approaches," *The Journals of Gerontology Series B: Psychological sciences and social sciences*, vol. 52, no. 5, pp. P197–P205, 1997.

[31] S. Cohen, "Social relationships and health," *American psychologist*, vol. 59, no. 8, p. 676, 2004.

[32] A. Rossi, K. Dautenhahn, K. L. Koay, and M. L. Walters, "Human perceptions of the severity of domestic robot errors," in *International Conference on Social Robotics*. Springer, 2017, pp. 647–656.

[33] S. 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.

[34] J. Laprie, "Dependable computing and fault tolerance: Concepts and terminology," in *Twenty-Fifth International Symposium on Fault-Tolerant Computing, Highlights from Twenty-Five Years*. IEEE, 1995, p. 2.

[35] R. Ross, R. Collier, and G. M. O'Hare, "Demonstrating social error recovery with AgentFactory," in *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 3*. IEEE Computer Society, 2004, pp. 1424–1425.

[36] J. Carlson and R. R. Murphy, "How UGVs physically fail in the field," *IEEE Transactions on robotics*, vol. 21, no. 3, pp. 423–437, 2005.

[37] G. Steinbauer, "A survey about faults of robots used in RoboCup," in *RoboCup 2012: robot soccer world cup XVI*. Springer, 2013, pp. 344–355.

[38] D. J. Brooks, "A human-centric approach to autonomous robot failures," Ph.D. dissertation, University of Massachusetts Lowell, 2017.

[39] M. Giuliani, N. Mirnig, G. Stollnberger, S. Stadler, R. Buchner, and M. Tscheligi, "Systematic analysis of video data from different human-robot interaction studies: A categorization of social signals during error situations," *Frontiers in psychology*, vol. 6, p. 931, 2015.

[40] A. Pandey and R. Gelin, "A mass-produced sociable humanoid robot: Pepper, the first machine of its kind," *IEEE Robotics & Automation Magazine*, no. 99, 2018.