

# How can humans and robots communicate better?

## Abstract

We propose a new approach for human-aware Artificial Intelligence (AI) systems and human augmentation based on the Johari Window theory. For effective and safe interaction, humans and AI systems, such as mobile robots, must share common goals, have a mutual understanding of each other, and know relevant aspects of each other's current states. The interaction between the human and the robot is also the mechanism for moving information between the Johari window panes associated with the human or robot. According to Johari Window Theory the goal of good communication is to expand the "Open area" square both horizontally and vertically. Vertical expansion occurs when we include information based on personalization, and horizontal expansion occurs via a structured trial-and-error process called reinforcement learning. For personalization, we propose to detect and recognize the humans' emotions and gestures, so the robot can respond accordingly. For reinforcement learning, we propose to develop a new model we call, multi-focus attention deep reinforcement learning, which is based on a control mechanism presented by Kahneman's Theory of two systems, or "Thinking Fast and Slow". If robots can "read" gestures of a human, this will determine the success of a task, and learning will occur through trial-and-error or reinforcement learning. For example, when the human becomes more aware of whether his/her facial expressions and gestures match the intent; this expands information vertically into the Johari window pane for the human. When the human's gestures and facial expressions become consistent with intent, the information will also appear in the open Johari window pane for the robot. In this way, trusted communication between humans and robots maximizes performance and safety.

**Keywords:** human -robot interaction, johari window model, robots, communication, behaviors

Volume 6 Issue 4 - 2020

Mariofanna Milanova,<sup>1</sup> Belinda Blevins-Knabe,<sup>2</sup> Lawrence O'Gorman<sup>3</sup>

<sup>1</sup>Computer Science Department, University of Arkansas at Little Rock, USA

<sup>2</sup>Department of Psychology, University of Arkansas at Little Rock, USA

<sup>3</sup>Nokia Bell Labs, USA

**Correspondence:** Mariofanna Milanova, Computer Science Department, University of Arkansas at Little Rock, USA, Tel +1 5015698681, Email [mgmilanova@ualr.edu](mailto:mgmilanova@ualr.edu)

**Received:** October 31, 2020 | **Published:** November 27, 2020

## Introduction

Currently, there is an increased focus on human-aware AI systems—goal-directed autonomous systems that are capable of effectively interacting, collaborating, and teaming with humans.<sup>1</sup> The research and development of human-like robots is growing. For example, at Amazon, robots play increasingly key roles in warehouses. They need to react in real-time to changes in their physical environment or the behavior of the people they are helping. Several drawbacks to the humanization of robots are presented in Robert LP, et al.<sup>2</sup> Robin Winsor at the end of his fascinating TED talk concluded that it is time to think about how we will guide the development of human-like robots.<sup>3</sup> The robot is our child and is our responsibility to create better communication between robots and humans. For effective and safe interaction, humans and AI systems, such as mobile robots, must share common goals, have a mutual understanding, and know relevant aspects of each other's current states.

Nonverbal behavior such as hand gestures, body positions, emotions, and head nods play a vital role in human communication.<sup>4</sup> During an interaction, the behavior of one person can be influenced by the behavior of another person. An example is when people mimic head nods in agreement. Sixty to 65 percent of interpersonal communication is conveyed via nonverbal behaviors.<sup>5</sup> Recent studies have brought new attention to how gesture communicates personality and emotion.<sup>6</sup> The study shows that small modifications of the gesture affect the interpretation of the meaning of the gesture.

We propose a new approach for designing human-aware AI systems including non-verbal behavior. Below are the three main processes we will follow in the development of the proposed architecture. First, we explain the communication between the robot and the human based on Johari Window for human-aware AI systems. Second, we explain how

the information flows through the four panes (quadrants). Third, we describe the new multi-focus attention deep reinforcement learning model as based on the control mechanism presented by Kahneman's theory of two systems.

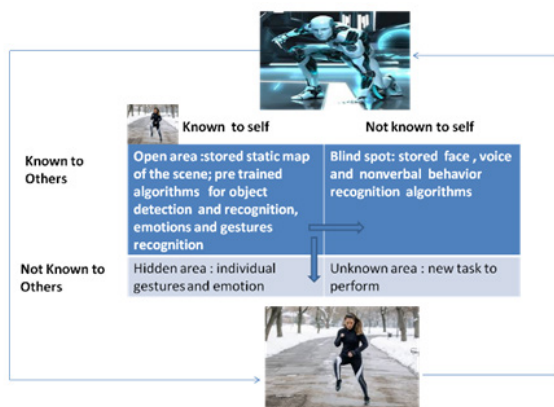
## Communication between robot and human based on Johari Window for human-aware AI systems

The Johari Window (Luft & Ingham, 1955) provides a model for describing the process of human interaction and understanding self-awareness.<sup>7</sup> This tool for personal development has been used to improve communication between individuals, within relationships, and within teams. The Johari Window divides self-awareness into four quadrants:

- i. Open Area: We know things about us and that others also know those things about us.
- ii. Hidden Area: We know things about us that others don't know about us.
- iii. Blind Area: We don't know things about us that others know about us.
- iv. Unknown Area: We don't know things about us and others also don't know those things about us.

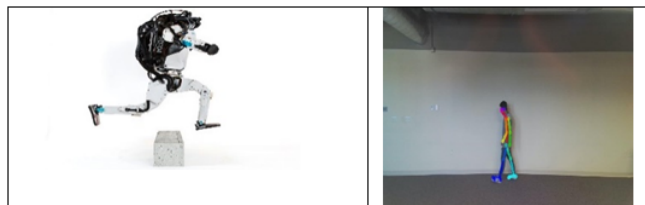
These four states can be applied to different types of communications and problem solutions. According to the Johari Window model the most important factor in the development of human relationships is disclosure. Through self-disclosure, boundaries are permeated, costs and rewards are measured, information flows through the four panes (quadrants), and understanding is developed as disclosers 'peel off'

the superficial layers to the core personality. The information included in the panes is unlimited. To succeed on a task or collaboration, humans and AI systems must share common goals and have a mutual understanding of them and relevant aspects of their current states. The open area is a pane accessible to both the human and robot with information known to both. It contains a static map of the scene and pre-trained algorithms for object detection and recognition, emotions, and gesture recognition algorithms. In the hidden area, private information that the individual may want to keep secret or may think is irrelevant is included. Robots may not keep secrets because they do not have self-awareness. However, they may have information they do not know is relevant. The blind spot stores algorithms necessary for perception, cognition, and behavior that operate below the level of the human’s explicit awareness. The unknown area represents new tasks that are in the future. Until a new task arrives we do not know what information humans and robots will need to work and understand each other. Figure 1 shows an example of a Johari Window for a robot interacting with a human. The task is for the robot to learn to run.



**Figure 1** Johari Window for human-aware AI system that is learning the task of running.

The task used for demonstration can be anything that the human-robot pair do together. As shown in Figure 2, an example is the task of the robot mimicking a jump performed by the human and learning actions from skeleton representations of the human body. Promising work on robotic mimicry has been done on facial expressions, speech and the imitation of gestures, all inherent components of human communication. Stolzenwald and Bremner created a system verifying the hypothesis that emotion and personality are recognizable in gestures and form the basis of a person’s gesturing style.<sup>8</sup>



**Figure 2** Robot mimicking a jump performed.

The interaction between the human and the robot is the mechanism for moving information between the Johari window panes associated with the human or robot. The arrows indicate the flow of information. The goal of good communication is to expand the “Open area” square both horizontally and vertically. The horizontal expansion is possible with reinforcement learning, while the vertical expansion is possible when we include information based on personalization.

## Personalization and achieving information flows through the four panes

Personalization helps the robot learn a new human partner’s unknown gestures, emotions, body position, and respond appropriately. In the beginning, progress is slow but with experience, the robot can adapt. Our team has considerable research in the area of analyzing nonverbal behavior,<sup>9</sup> recognizing the emotion of children with autism,<sup>10</sup> and real-time intention recognition.<sup>11</sup> Currently, we are working on multimodal health monitoring architecture for companion robots.<sup>12</sup> We proposed to develop a customized Deep Learning model to recognize individual human gestures, emotions, and voice. We already have a system prototype monitoring gait of people from the video as well as classifying normal from abnormal gestures and body posture using skeleton representations of the human body. New Transfer Deep Learning models are used at the starting point, and after that fine-tuning is performed on the model to incorporate an individual’s gestures and joints movements.

## New multi-focus attention deep reinforcement learning model based on control mechanism presented by Kahneman’s theory of two systems

When we include non-verbal behavior in human–robot interactions the question arises of whether robots and autonomous intelligent systems need to model the mental states of their human partners. To answer this question most researchers study trust and its effects on teamwork between autonomous agents and their human teammates and “theory of mind”. Theory of mind research examines the naïve psychology that humans use to understand their own behaviors and those of others.<sup>13–15</sup>

There are risks of applying the theory of mind to programming robots. The main question is: Do robots need a theory of mind?<sup>16</sup> An extensive review of 40 years in cognitive architecture research is presented in Kotscheruba, et al., 2020.<sup>17</sup>

According to Rivers’ research, the control mechanism should be based on Kahneman’s theory.<sup>18</sup> According to Kahneman’s theory System I (fast, with fewer inhibitory controls) creates a coherent pattern of activated ideas in associative memory. System I is currently presented in many neural network models. An example is the Hopfield neural network. Neural nets have already been used to create algorithms that demonstrate some aspects of the Theory of Mind.<sup>19,20</sup> The modeling of System II (slower and reflective) is a challenging problem. Deep Reinforcement Learning is the logical solution if the problem is well defined. There are several humanoid robots capable of mimicking humans.<sup>21,22</sup>

A possible task is the robot mimicking a jump performed by the human and learning actions from skeleton representations of the human body. In multi-agent reinforcement learning tasks, agents perceive partially observed states, which are part of the environment state. Agents have to communicate with each other to share information and cooperate to solve the task. Our hypotheses are:

- i. Humans can focus on multimodal information simultaneously. To demonstrate this, we include in our model a parallel multi-attention mechanism for fast and efficient learning.
- ii. Humans use System I to detect objects and extract features based on previous experience, then use System II to determine what information is relevant to solving the task.

Our proposed Deep Q learning network incorporates multi-focus attention to coordinate two modules: the first module estimates state-action value using data from the position and poses of the robot and position and type of objects in the scene; the second module is used for communication between the robot and human. Using parallel attention, the model can learn not only the policy for correct movements of the robot and the human but also the communication protocol between them.

By the human: skeleton-based human action recognition involves predicting actions from skeleton representations of human bodies.

Two attention mechanisms are working in each partial state in the learning process:

- a. One is tracking locations and trajectories of the moving robot and moving human and
- b. Another is tracing joints and movements of the human.

In each partial state in RL the combined feature vector is created and State-Action Value for state  $Q_t$  is estimated. The system will be implemented using the Robot Operating System ROS on Jetson TX2, or Jetson Nano platform.

## Conclusion

In this work we claim that for better human-robot communication a fruitful approach might be to consider team building activities with less pain and more gain using Johari Window model.

## Funding

The research was sponsored by NOKIA Corporate University Donation number NSN FI (85) 1198342 MCA and National Science Foundation under Award No. OIA-1946391.

## Acknowledgements

University of Arkansas at Little Rock, USA and Nokia Bell Labs, Murray Hill, New Jersey, USA.

## Conflicts of interest

The authors declare that there was no conflict of interest.

## References

1. Subbarao K. Challenges of Human –Aware AI Systems. *Artificial intelligence*. 2019.
2. Robert LP. The growing Problem of Humanizing Robots. *IRATJ*. 2017;3(1):247–248.
3. <https://www.youtube.com/watch?v=f7dhOHMX0js>
4. Fatemeh Noroozi, Adrian Corneanu C, Tomasz Sapiński, et al. Survey on Emotional Body Gesture Recognition. *IEEE Transactions on Affective Computing*. 2018;1–19.
5. Burgoon JK, Guerrero LK, Floyd K. Nonverbal Communication. USA: Allyn and Bacon; 2010.
6. Gabriel Gastillo, Michael Neff. What do we express without knowing? Emotion in Gesture. *AAMAS*. 2019;13–17:702–710.
7. <https://www.communicationtheory.org/the-johari-window-model/>
8. Stolzenwald J, Bremner P. Gesture mimicry in social human-robot interaction. IEEE International Symposium on Robot and Human Interactive Communication. 2017:430–436.
9. Mariofanna M, Leonardo B. Video-Based Human Motion Estimation System. *HCI*. 2009;11:132–139.
10. Anwar S, Mariofanna M. Real time Face Expression Recognition of Children with Autism. *IAEMR journal*. 2016:1–7
11. Suzan Anwar, Mariofanna Milanova, Andrea Bigazzi, et al. Real time intention recognition. *IECON*. 2016:1021–1024.
12. Xinyi Liu, Imran Sarker, Mariofanna Milanova, et al. Video-Based Monitoring and Analytics of Human Gait for Companion Robot, Proceedings of International Workshop. *Smart Innovation, Systems and Technologies*. 2020.
13. Beaudoin C, Leblanc E, Gagner C, et al. Systematic review and inventory of theory of mind measures for young children. *Frontiers in Psychology*. 2020.
14. Derksen DG, Hunsche MC, Giroux ME, et al. A systematic review of theory of mind's precursors and functions. *Zeitschrift Für Psychologie*. 2018;226(2):87–97.
15. Henry JD, Phillips LH, Ruffman T, et al. A meta-analytic review of age differences in theory of mind. *Psychology and Aging*. 2013;28(3):826–839.
16. Rod Rivers. Do Robots Need Theory of Mind?. 2020.
17. Kotseruba I, Tsotsos JK. 40 years of cognitive architectures: core cognitive abilities and practical applications. *Artif Intell Rev*. 2020;53:17–94.
18. Daniel Kahneman. Thinking, Fast and Slow. 2013.
19. Rabinowitz NC, Perbet F, Song HF, et al. Machine Theory of mind. *Artificial intelligence*. 2018;10:6723–6738.
20. Milanova M, Bükler U. Object recognition in image sequences with cellular neural networks. *Neurocomputing*. 2000;31(1-4):125–141.
21. Janis Stolzenwald, Paul Bremner. Gesture Mimicry in Social Human-Robot Interaction. *IEEE RO-MAN*. 2017:1–7.
22. Yoon Y, et al. Robots Learn Social Skills: End-to-End Learning of Co-Speech Gesture Generation for Humanoid Robots. *ICRA*. 2019:1–7.