2023 32nd IEEE International Conference on Robot and
Human Interactive Communication (RO-MAN)
August 28-31, 2023. Paradise Hotel, Busan, Korea

2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) | 979-8-3503-3670-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/RO-MAN57019.2023.10309527

Communication as Joint Prediction: A Case Study of Robot-Mediated Pretend Play with Children at a Kindergarten

Hideki Kozima[1]

***Abstract*— Children naturally engage in pretend play with others in their daily life. The children share each other's expectations toward objects and people around them. Such expectation-sharing can be considered the fundamental process of social communication, where we exchange beliefs, desires, and intentions. Firstly, the present paper examines some cases of such expectation-sharing in pretend play observed in our longitudinal robot-mediated interaction with 27 preschoolers (3-to-4-year-olds) at a kindergarten. The robot, which the researchers in another room remotely controlled, has a simple appearance and motion to accept various expectations from the children. The robot functioned as a pivot for the children to exchange each other's expectations towards the robot, toys, and people. The children might bring expectations incongruent with each other, but such expectations gradually converged into a consistent one, on which they understood each other's verbal and non-verbal actions. Secondly, this paper interprets and models these cases using the predictive coding theory. According to the theory, the children project their predictions onto the shared environment and update the predictions by minimizing the error against reality. This act of projection, which was performed as labeling (e.g., a block as a rice ball) or pretend action (e.g., feeding it to the robot) by the children, is a form of active inference to modify the environment to match the prediction. The children exchanged and coordinated their predictions through visible action collectively accumulated in the playing environment around the robot. We conclude that joint prediction is the fundamental process of social communication, where we coordinate each other's behavior for collaboration.**

I. INTRODUCTION

We can imagine things that do not exist or have not happened before our eyes. We can even imagine impossible events or completely new things. This ability of imagination and creativity is probably unique to humans, which has given us the ability to produce stories and tools [1]. Even preschool children can imagine such nonexistent things and events in their pretend play in their daily life [2], [3], [4]. For example, it is natural for children to play house, where each has a particular role (e.g., mother), and their toys have imaginary meanings (e.g., a teddy bear for their baby, a block for a piece of food to feed).

This ability for pretense is not only for creating fantasy but also for communication among children [5], [6]. Children collectively engage in pretend play, where they refer to and exchange each other's imagination toward the visible scene in front of them. Each of the children may have different imagination at the beginning; however, they gradually converge into a shared set of expectations on which they understand each other's verbal and non-verbal actions. This convergence of expectations is the fundamental process of everyday communication, where we imagine and exchange each other's beliefs, desires, and intentions in order to collaborate on shared goals.

We carried out a longitudinal series of children-robot interactions at a kindergarten, where we observed such expectation-sharing in their pretend play involving our robot. Those children interacted with each other and resonated their imagination toward the robot. The children shared their expectations of the robot's condition (e.g., being sad or sick) and its role in the play (e.g., a baby or a customer) and collaboratively developed a storyline. As they collectively overlap and coordinate their verbal and non-verbal actions in the shared environment of pretend play, the children's internal fantasies converge within a specific range, consistent with their major predictions and actual experiences. We have observed a lot of such scenes during our one-year-long observation of children's play in the 3-year-olds' class.

This paper interprets this phenomenon from the perspective of a cognitive model called "predictive coding theory" [7], which has been the focus of much attention in recent years in cognitive science. In the predictive coding theory, our brain makes predictions (or hypotheses) on incomplete sensory information. By Bayesian updating of the predictions to eliminate the error against the sensory information, we can estimate the hidden state (unobservable information) behind the sensory information. Also, the sensory information can be modified by acting on the environment to improve its consistency with the predictions. By the prediction updating process (perception) inside and the environment updating process (action) outside in parallel, one can maintain the alignment between the internal model and the external environment, increasing capability and adaptability to the environment.

The following section introduces the robot we used in our longitudinal observation at the kindergarten. Section III describes typical cases of children's pretend play involving the robot. Section IV interprets and models these cases using the predictive coding theory. The final section discusses the ecological and theoretical validity of the present work and puts it in a future perspective.

II. ROBOT AT A KINDERGARTEN

The robot we used in our observation at the kindergarten is "Keepon" [8], a simple creature-like robot specially designed for non-verbal interactions with preschool children. We describe here the design and function of the robot, how

[1]Graduate School of Education, Tohoku University, Kawauchi 27-1, Aoba, Sendai, Miyagi 980-8576, Japan xkozima@tohoku.ac.jp

Fig. 1. Keepon's internal structure and external appearance.

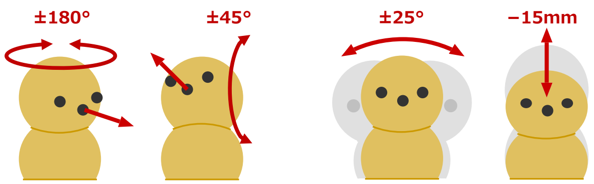

Fig. 2. Keepon's movement for expressing attention and emotions.

it is tele-controlled by the researchers, and how we situate it in the 3-year-olds' class at the kindergarten.

A. Keepon, the Robot for Children

Keepon is originally developed for research and therapeutic practices for autistic children [9]. As illustrated in Fig. 1, the yellow body (120mm tall) of silicon rubber has two video cameras for the eyes and a microphone at the nose. The internal mechanism, which has 4 degrees of freedom driven by DC motors with planetary gears and rotary encoders for quiet and life-like motion, enables the body to express its attention (by gaze/face direction) and emotion (by rocking and bobbing the body), as illustrated in Fig. 2. Note that most of the driving mechanism is enclosed in the lower cylindrical case to ensure the safety of children interacting physically with it.

Keepon's appearance and function stem from the principle of "minimal design" [8], in which we restrict the amount of information it emits through its posture and movement. For autistic children, Keepon's minimal design enables them to understand the robot's attentional and emotional states without causing sensory information overload [9], [10]. For typically-developing children, which we focus on here, Keepon's minimal design allows them to project various meanings on it. Its face is not expressive but accepts children's various expectations, such as laughing and crying.

B. Remotely Controlling the Robot

We installed Keepon in a conical casing and covered it with cushions for children's safety, as illustrated in Fig. 3. We also put batteries and wireless equipment in it. Keepon can autonomously interact with children using visual information processing of human faces and body motions; however, in our participant observation at the kindergarten, we remotely control the robot in real-time from another room. The primary reason is that current AI technology does not allow us to produce socially meaningful responses, such as the pretense of eating (by bobbing motions) when a child tries to feed something to it.

Fig. 3. Keepon with batteries and wireless equipment.

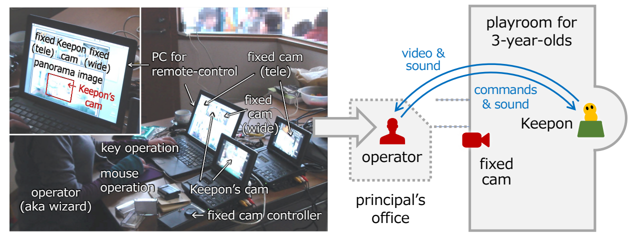

Fig. 4. The setting for remotely controlling the robot at the kindergarten.

Fig. 4 illustrates the remote control set at the kindergarten. We stayed in the principal's office, where we remote-controlled the robot in the children's playroom with the help of video images from Keepon's eyes and a fixed pan-tilt camera installed on the ceiling of the playroom. A wireless communication system establishes the video, sound, and data links between the operator and playroom sides. The operator controls the direction of Keepon's head by mouse clicks for ballistic motions or mouse drags for smooth pursuit. Keystrokes produce bobbing and rocking movements with artificial sounds, such as "pong-pong-pong", output from Keepon's onboard speaker. Children in the playroom never see the operator or the tele-controlling system throughout the observation.

C. Interactions Situated in the Playroom

The human operator consistently remote-controlled Keepon in the playroom using the graphical interface shown in Fig. 5. The upper layer shows the images from (1) Keepon's onboard camera, (2) the fixed camera for a close-up view for the operator to understand what is happening around Keepon, and (3) another fixed camera for a wider angle of view for figuring out what is happening in the playroom. The lower part of the interface shows a 360-degree panorama view from Keepon's viewpoint. Only the part corresponding to Keepon's gaze direction updates in real-time; other surrounding parts are a patchwork of still images most recently taken by Keepon's camera when it directs its gaze toward the directions.

This controlling interface allows the operator to produce

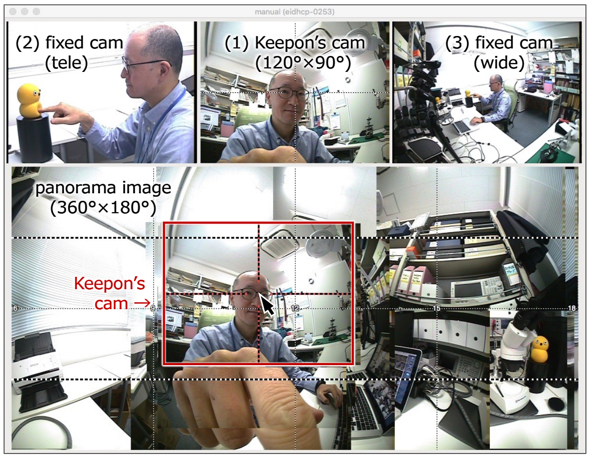

Fig. 5. The graphical interface for situated interaction in the playroom. (This example is taken in the author's lab for the explanation.)

socially situated behavior in the company with the children. Usually, several children in a group surround Keepon simultaneously at the kindergarten, in contrast to the cases with autistic children, who typically interact with Keepon one by one. Being surrounded by the children, the operator can selectively engage with one of the children at a time, then turn to another child to engage with. Thus, the children sometimes directly act on Keepon and sometimes observe others acting on Keepon, having a chance to compare themselves and others.

III. CASE STUDIES

We introduced Keepon to a class of 27 children (12 boys and 15 girls) and three teachers at the kindergarten. Those children were 36 to 47 months old at the beginning of April when schools and kindergartens begin in Japan. The reason why we especially focused on this age (3 to 4 years old) is that we wanted to investigate the emergence of the "Theory of Mind" [11], [12], which would play a crucial role in their collective pretend play. We carried out 25 series of longitudinal observations of children-robot interaction in this class for 12 months from April to March, which means that those children's average age during our observation was around 4.0 years.

Our observation focused on the morning time (8:30 to 11:30 a.m.). The children freely played in their playroom in the first half (8:30 to 10:00). In the latter half (10:00 to 11:30), they engaged in group activities organized by the teachers, as shown in Fig. 6. We obtained consent from the parents before the observation. However, the children did not know if someone was remotely operating Keepon.

We here describe some of the representative cases we identified as examples of joint prediction to co-create fantasies among the children. Note that we here processed the pictures for privacy protection; we translated the children's utterances from Japanese and transcribed them using a standard notation in conversation analysis [13].

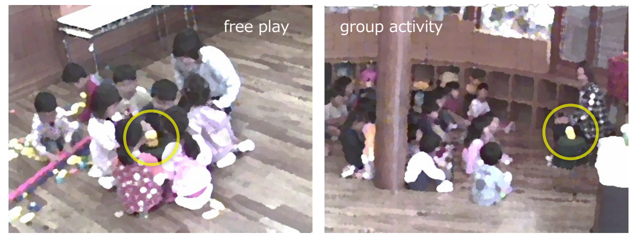

Fig. 6. Keepon in the children's free play and group activity.

Fig. 7. Case 1: Children give Keepon a salmon rice ball.

A. Case 1: Keepon and a Salmon Rice Ball

The first case, illustrated in Fig. 7, shows how children's pretense on an object was elaborated through verbal and non-verbal interactions among the children and a teacher.

M_1,R ((Fed a toy block to Keepon together.))
(1) K ((Cutting in and inviting a teacher.)) *Look! Here is Keepon!*
M_1 ((Facing the teacher.)) *Look. He is eating! Look. He is eating!*
Tchr *Oh! What is he eating?* ((Sat down in the children's circle.))
$*$ M_1 ((Smiled to the teacher but cannot answer her.))
K ((Feeding a block to Keepon.)) *Have this rice ball!*
Tchr *A rice ball. What kind of rice ball?*
K ((Facing the teacher.)) *A salmon rice ball!*
(2) Tchr ((Giving a clap and smiling.)) *Wow! My favorite salmon rice ball.*
(3) K *Here it is.* ((Gave the block to the teacher.))
Tchr ((Pretending to eat it.)) *This tastes really good.*
K ((Fed the same block to Keepon.))
R *Wow! Eat it! Eat it! Eat the salmon rice ball!*
K *He is eating it, eating it!*
((Facing the teacher.)) *Keepon ate it.*
Tchr *Keepon ate it. You did it well!*
(4) M_1 ((Facing the teacher.)) *He said it was delicious.*
Tchr *Yes, he said delicious.*

Here we see that M_1 did not have a particular assumption of what food she was feeding Keepon ($*$). However, as the interaction between K and the teacher unfolded, those

Fig. 8. Case 2: Children take care of Keepon's head injury. (Note that (3) and (4) were taken from Keepon's on-board camera.)

Fig. 9. Case 3: A child seeks a company to co-create a fantasy.

children's assumptions gradually got detailed into "a rice ball" and "a salmon rice ball". As a result, M_1's assumption also got specified (4). The teacher's intervention worked well in elaborating a collective expectation among children. She asked the children step by step what they expected in Keepon and the blocks, inducing their spontaneous imagination and coordination.

B. Case 2: Keepon Needs Medication

The second case, illustrated in Fig. 8, shows how the children's caretaking behavior toward Keepon gradually expanded and sophisticated as their verbal and non-verbal actions accumulated around Keepon.

(1) R,T ((Put their weight on Keepon, making a crack in its head. They found an internal part sticking out of the head, and awkwardly left the scene.))

(2) K,M_2 ((Reported Keepon's injury to a teacher, saying "an egg" and "a bump". The children around Keepon mentioned "poor" and "painful".))

(3) K ((Showing an empty cup to Keepon.)) *How about medicine? Need one?*

K ((Keepon noded, then K spooned the medicine for Keepon several times.))

$*$ A_1 ((Watching K's behavior, A_1 tried to give a toy dish to Keepon.)) *This is Keepon's dish...*

Tchr ((Talking to K.)) *Does Keepon take medicine?*

K ((Feeding Keepon.)) *Yeah, because it makes his head scar better.*

$*$ A_1 ((Imitating K's behavior, A_1 reached out a spoon toward Keepon.))

K *Bitter?* ((Keepon shook its head.)) *Was that OK?* ((Keepon nodded.)) *Now you need some water. Water, water...* ((K left Keepon for water.))

K ((Returning to Keepon shortly and feeding water to Keepon.)) *It was not bitter.* ((Feeding some more water.)) *Now you will be fine!*

(4) M_2 ((Having observed K's behavior, brought an empty cup to Keepon.)) *This is milk. Let's give this milk to Keepon.*

$**$ Tchr *Is that warm milk?*

M ((After thinking for a while.)) *It's cold.*

$**$ Tchr *Cold? Warm milk would be better.*

M_2 ((After thinking.)) *Now it has got warmer!*

K,M_2 ((Feeding Keepon the warm milk together.))

In the episode above, K was the first child to take care of Keepon by giving it medicine and water (3). Observing that, A_1 tried to imitate K's behavior superficially ($*$), but M_2 understood the meaning of K's action and extended it to giving milk to Keepon, first cold and then warm (4). The teacher's interventions ($**$) helped M_2 understand what was especially expected in the shared situation.

C. Case 3: Pretend Play Needs Company

The third case, illustrated in Fig. 9, exemplifies that children's pretend play sometimes breaks down as the members leave or join. In the following episode, a child seeks company to play with after missing her company in the pretend play.

R,T ((Feed Keepon using cups and spoons. T was worrying about the loosening of Keepon's eye.))

(1) R ((Looking into Keepon's eyes.)) *Poor, Keepon...*

T ((To R.)) *He is sick. He is coughing.*

(2) R ((Looking up at the distance.)) *Coughing...*

T ((Stroking Keepon.)) *He is coughing. It's a pity.*

T ((After a while, R left Keepon. T cared for Keepon alone, but T was looking around for others.))

(3) T ((Showing up her spoon for feeding.)) *Does anybody want this?*

T ((Nobody responded. T threw the spoon away.))

(4) T ((After a while, still not having any company, T moved Keepon by pushing it toward L and A.)) *Here comes Keepon! His name is 'Keepon.'*

L ((Looking into the eyes.)) *His eyes are loosening...* ((L, A, and T interacted with each other.))

Here we see that pretend play often needs company. Children have the motivation to share their expectations with others. Having had a joint engagement, T missed her playmate R. Then, T tried hard to invite another playmate but failed. So, T pushed Keepon to slide it on the hardwood

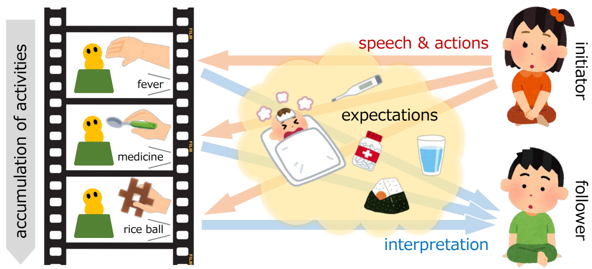

Fig. 10. Initiation and sharing of expectations in pretend play.

floor toward other children nearby, with whom T restarted unfolding her fantasy to share.

IV. JOINT PREDICTION

In the previous section, we examined some cases of children's pretend play, exemplifying how fantasies are collectively formed. We here interpret this emergent process of shared fantasies using the predictive coding theory [7] and extend this idea to the fundamental model of social communication.

A. Co-Creation of Fantasy

As we witnessed in these cases, children individually presume their own imaginations, which weakly determines the roles of the robot, other children involved, and the objects in the environment. As illustrated in Fig. 10, a child who initiates the pretend play performs verbal and non-verbal actions based on the presumption. Such activity accumulates around the robot, becoming visible and traceable for the followers to interpret and share the expectation.

Fantasy creation thus starts from an individual accumulation of pretend actions, but having playmates, extends later into a collaborative one [5]. At the beginning of the collective interactions, each of the children may have different presumptions inconsistent with the others. However, as they interpret other's activities and perform meaningfully to the existing accumulation of the activities, their expectations gradually converge into a consistent one. Thus, the follower becomes a new participant in the ongoing pretend play to contribute to further accumulating the activities. We call such a collectively formed set of consistent expectations "shared fantasy."

B. Predictive Information Processing in Pretend Play

We assume that the predictive coding theory [7] can explain the underlying mechanism for the collective construction of shared fantasies. When we perceive something out in the world, we project our expectation toward incomplete and ambiguous sensory information coming from the outside, as illustrated in Fig. 11. Then, our brain computes the predictive error, which is the difference between the top-down predictive expectation and the bottom-up sensory evidence, which guides us to update the expectation using Bayesian inference, where the original prediction is the prior belief, the incoming sensation is the likelihood, and the updated prediction is the posterior belief about the state of

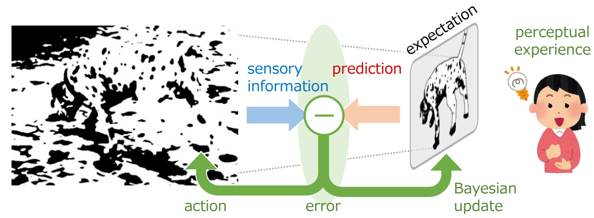

Fig. 11. Predictive process for generating perceptual experiences.

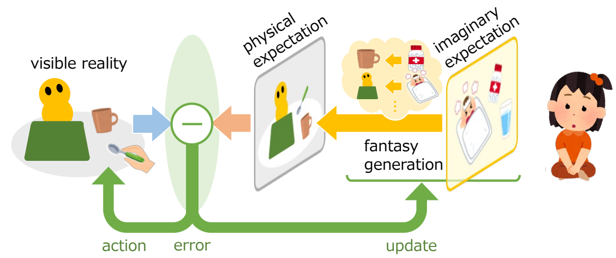

Fig. 12. Predictive process in counterfactual pretend play.

the outside world. We also act on the world, such as getting closer to the object, to change the incoming sensory evidence so as to minimize the error against the prediction.

We apply this predictive information processing model to the cognitive process underlying pretend play, as illustrated in Fig. 12. The left half of the figure shows the same process as that in Fig. 11, which produces a low-level primary perception of the real world per se. The right half is an additional generative process that translates the imaginary expectation into a physical one that can be directly compared with reality. The predictive error will then be used for updating the fantastic world, namely the imaginary expectation and how it translates into a physical one. In addition, the predictive error can also be decreased by changing the real world through verbal and non-verbal actions.

C. Collective Creation of Shared Fantasy

Children collectively engage in pretend play. As new members join in and the story unfolds, the children's fantasy expands and sometimes collapses autonomously. As illus-

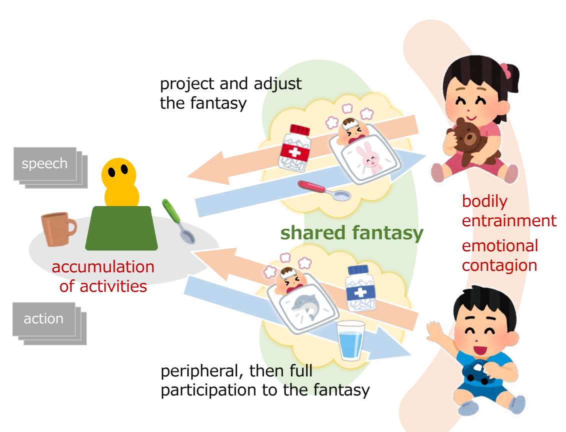

Fig. 13. Collective creation of and engagement in pretend play.

trated in Fig. 13, a child transforms her expectations into visible activities through projection and adjustment, accumulating them in the real world. When another child joins in, he estimates the invisible meaning of the robot, objects, and other children, hidden in the accumulated activities. This process is a kind of "peripheral participation" [14], in which the newcomer tries to adapt to the existing imaginary setting; however, in the later stage, it develops into "full participation" [14], in which the newcomer modifies and expands the fantasy collaboratively with others.

The children's engagement in pretend play anchors in the following two sources of information, as illustrated in Fig. 13 again. One is, as mentioned above, the visible accumulation of activities, such as speech and actions the children have performed according to their expectations. The other is the entrainment of their bodily motions and the contagious resonance of emotional responses. Being anchored to the world's objective state and the children's resonating subjective states, a "shared fantasy" emerges as the common field for exchanging perceptions and actions among the children.

V. CONCLUSIONS

In this paper, we examined some cases of expectation-sharing in pretend play observed in our robot-mediated interactions with the preschoolers at the kindergarten; then, we modeled the emergence of "shared fantasy" in their pretend play using the predictive coding theory. In our model, children project their imaginary expectations for pretend play onto the robot and the environment through verbal and non-verbal actions. Those actions accumulate around the robot and become visible to other children, enabling them to perform the predictive processing to estimate and elaborate their collective expectations.

Some studies suggest that our perception is a "controlled hallucination" [15], which implies that our prediction-based inference generates our perceptual experiences. In our cases of pretend play, the children collectively generated, exchanged, and coordinated their counterfactual experiences. Being anchored on the visible accumulation of their verbal and non-verbal actions and the resonance of their bodily motion and emotional responses, the children collectively produced "shared fantasy" in the same way as "controlled hallucination."

Based on our model of the emergence of "shared fantasy," we can extend the idea to a broader model of social communication. In our daily life, we understand others' mental states, such as beliefs, desires, and intentions, and exchange them with each other for collaboration and for having a sense of unity [16]. Such a social ability, namely the "Theory of Mind" [11], [12], is the core of human-specific traits, but how it emerges and develops in children is still not known in detail. One of the reasons is that most past studies dealt mainly with an individual rather than collective activities in the company of children. Our model of the emergence of "shared fantasy" would explain how the intersubjective understanding of others' expectations emerges from collaborative activities in the company of children. We will deeply investigate this broader model in our further research.

To facilitate the emergence of "shared fantasy" in children, simple robots like Keepon works well as a pivot for their exchange of speech and actions [8]. For example, Keepon can selectively engage with one child through eye contact, which encourages others to observe how she acts on it. This functionality enables them to switch between an actor and an observer, to compare one's expectation with another's, and to co-create the "shared fantasy" to engage collectively.

We here focused on toddlers' social development; however, the use of robots also applies to the wider field of robot-assisted learning and education [17], which includes assisting older children's academic and social achievement. Even in such support, collective activities and sharing imagination are fundamental to developing children. We will further explore the advantages of using robots to foster children's collective imagination.

ACKNOWLEDGMENT

The author is deeply grateful to the children, their parents, and the teachers at the kindergarten, also to Dr. Cocoro Nakagawa and Dr. Marek P. Michalowski, who helped with the robot-mediated participant observation.

REFERENCES

[1] P. K. Smith, Social and Pretend Play in Children, in A. D. Pellegrini, and P. K. Smith (eds.), *The Nature of Play: Great Apes and Humans*, The Guilford Press, pp.173–209, 2005.
[2] J. Piaget, *Play, Dreams and Imitation in Childhood*, Routledge, 1951.
[3] C. Howes, Sharing fantasy: social pretend play in toddlers, *Child Development*, vol.56, pp.1253–1258, 1985.
[4] D. G. Singer and J. L. Singer, *The House of Make-Believe: Children's Play and the Developing Imagination*, Harvard University Press, 1990.
[5] A. S. Lillard, Pretend play skills and the child's theory of mind, *Child Development*, vol.64, pp.348–371, 1993.
[6] D. Bergen, The role of pretend play in children's cognitive development, *Early Childhood Research and Practice*, vol.4, 2002.
[7] K. Friston, The free-energy principle: a unified brain theory?, *Nature Reviews Neuroscience*, vol.11, pp.127–138, 2010.
[8] H. Kozima, M. P. Michalowski and C. Nakagawa, Keepon: a playful robot for research, therapy, and entertainment, *International Journal of Social Robotics*, vol.1, pp.3–18, 2009.
[9] H. Kozima, C. Nakagawa and Y. Yasuda, Children-robot interaction: a pilot study in autism therapy, *Progress in Brain Research*, vol.164, pp.385–400, 2007.
[10] B. Scassellati, H. Admoni and M. Matarić, Robots for use in autism research, *Annual Review of Biomedical Engineering*, vol.14, pp.275–294, 2012.
[11] D. Premack and G. Woodruff, Does the chimpanzee have a theory of mind?, *Behavioral and Brain Sciences*, vol.1, pp.515–526, 1978.
[12] A. M. Leslie, O. Friedman and T. P. German, Core mechanisms in 'theory of mind', *Trends in Cognitive Science*, vol.8, pp.528–533, 2004.
[13] J. Sidnell, *Conversation Analysis: An Introduction*, John Wiley & Sons, 2011.
[14] J. Lave and E. Wenger, *Situated Learning: Legitimate Peripheral Participation*, Cambridge University Press, 1991.
[15] A. Seth, *Being You: A New Science of Consciousness*, Dutton, 2021.
[16] M. Tomasello, M. Carpenter, J. Call, T. Behne and H. Moll, Understanding and sharing intentions: the origins of cultural cognition, *Behavioral and Brain Sciences*, vol.28, pp.675–735, 2005.
[17] A. Augello, L. Daniela, M. Gentile D. Ifenthaler and G. Pilato, Editorial: robot-assisted learning and education. *Frontiers in Robotics and AI*, vol.7, 591319, 2020.