25TH ANNIVERSARY VOLUME

A FAUSTIAN EXCHANGE: WHAT IS TO BE HUMAN IN THE ERA OF UBIQUITOUS TECHNOLOGY?

Toward mutual dependency between empathy and technology

Toyoaki Nishida

Abstract Technology explosion induced by information explosion will eventually change artifacts into intelligent autonomous agents consisting of surrogates and mediators from which humans can receive services without special training. Four potential problems might arise as a result of the paradigm shift: technology abuse, responsibility flaw, moral in crisis, and overdependence on artifacts. Although the first and second might be resolved in principle by introduction of public mediators, the rest seems beyond technical solution. Under the circumstances, a reasonable goal might be to create a mutual dependency between empathy and technology: Using technology to help people cultivate empathy among people, so empathy in the society may allow people to help each other to recover from disasters beyond the scope of the assumption underlying the society of mediators. I highlight an immersive collaborative interaction environment for helping people share first-person view and its application to building empathic agents.

Keywords Technology abuse · Responsibility flaw · Moral in crisis · Overdependence on artifacts · Empathy · Sharing hypothesis · Empathic agents · Teaching by indwelling · Learning by mimicking

T. Nishida (✉)
Kyoto University, Sakyo-ku,
Kyoto 606-8501, Japan
e-mail: nishida@i.kyoto-u.ac.jp

1 Info- and techno-plosion

The digital universe is exponentially expanding. The volume of information on the net has exhibited an exponential increase in the last decade. This phenomenon is often called information explosion or info-plosion (Kitsuregawa and Nishida 2010). In the info-plosion age, the user is becoming more interested in information services brought about by artifacts, rather than artifacts themselves. Info-plosion not only visualizes a long tail of desire that might serve as a driving-force for technology but also popularizes methods and tools for realizing desire. The collective intelligence empowered by the information infrastructure enables agile fabrication of a solution to desire. The more desire is satisfied, the more desire is generated.

Eventually, info-plosion will result in techno-plosion, causing a paradigm shift from the artifacts-as-designed-tools to the artifacts-as-autonomous-agents. In this paradigm, we live with autonomous agents for assistance in the living space. They will serve as an integrated interface between the real and cyber world. They will be able to collect the situational information around themselves through sensors and provide a suitable service with the user through actuators of various kinds. They will hang around you and help you both reactively and proactively. As the technology will realize the details, the human will not have to be a computer geek in order to benefit from the advanced information and service environment brought by intelligent autonomous agents. From the moral points of view, the advent of autonomous agents may more conform to humanity, for they will allow us to concentrate on the *end*, as opposed to the *means*, and ultimately, we will not have to pay much attention on the tools, or the means for the end, as autonomous agents will take care of all the details concerning the tools on behalf of us.

Unfortunately, there will be four major problems underlying the naive implementation of intelligent autonomous agents. The first problem is *technology abuse*. New technologies can be applied to illegal or malicious activities. Artifacts that merely act on behalf of the owner extend both the good and evil wills of the owner. The advanced technology might enable an intruder to sneak into your proximity without being noticed or permit a fraud to deceive you in a novel way. The past experiences and knowledge might not be able to prevent damages from happening. Serious problems may arise before people have become aware of them to take precautions (Nishida 2007).

The second problem is *responsibility flaw*. Perrow pointed out the difficulty of sustaining sensibility against accidents as an organization, and accidents might happen even normally (Perrow 1984). The more complex artifacts become, the less likely the humans can put them under control. Neither the product maker nor the owner of a complex artifact may take a full responsibility of an artifact, if it is fairly complex; the owner of the artifact cannot envision all possible outcomes caused by the artifacts she/he possess, for it might be simply beyond her/his intellectual capabilities; a product maker may not be able to make sure that the product may work properly under any usage scenarios.

The third problem is *moral in crisis*. In a complex and changing world, ethical conflicts might be caused when people have inconsistent beliefs in how the ethical principles are instantiated as moral rules, although people might know that ethical principles (Kant 1788; Rawls 1999; Decker 2008) are designed to protect the autonomy and dignity of humans and construe it as entailing the norm of protecting the weak from being used solely as a means to achieve an end by the strong. People might not be fully aware of the consequence of her/his dominance over the weak (Nishida 2009a).

The fourth problem is *overdependence on artifacts*. Introduction of autonomous agent will cause heavy dependence on artifacts. Individuals might use artifacts without judgment. The society might assume the infallibility of artifacts without rationale. There are strong concerns about heavy dependence on artifacts. Among others, Cooley exhibited a similar concern using "from judgment to calculation" (Cooley 2007) and gave a caveat of being overly depending on calculation rather than judgment. The heavy dependence will entirely remove motivations of thinking and imagination at the individual level and might bring about "empty brains" (Maurer 2007). Maurer warns that a serious break down of the computerized social infrastructure might cause a catastrophic disaster and bring the human society back to the Stone Age.

2 Computationally mediated society

The technology abuse and responsibility flaw problem might be resolved by introduction of a computational framework of mediating social interactions, in which public mediators are introduced in addition to surrogates and private mediators to bring about fair distribution of resources.

The design, implementation, and operation of public mediators are made by a public organization authorized in the society. Public mediators shall be built under a publicly transparent process and operated under public monitoring. Risks caused by public mediators may be underwritten by the public; even if the user may suffer from damage, it is reasonable that the public sector representing the community involving the user takes a responsibility to compensate for the damages caused by the mediators at a public cost. In contrast, private mediators may be employed at occasions when special conveniences are desired at the user's own cost.

This new framework is deemed as the world supported by "super intelligence" in which each human will not directly interact with each other or play social functions anymore; instead, people interact with each other through the public or private mediators. Inter-human interaction is realized by coupling the human-agent communication between the owner and her/his surrogate, and the social communication among mediators. Each person's intention is communicated to her/his surrogate in the human-agent communication. Surrogates interact with each other on behalf of the owner. Each surrogate attempts to maximize the satisfaction of the owner's intention, based on a specialized knowledge about how to plug-in to the network of mediators. In contrast, public mediators are fabricated and maintained by the public sector. Public mediators may attempt to maximize the harmony, if not the merit, of the society as a whole, by maximally arbitrating the requests and offering from participating agents, while private mediators may work for the owner or institution to which it belongs.

The computational framework of mediating social interactions appears to solve the technology abuse and the responsibility flaw problems. The technology abuse problem will be almost solved if not complete, as people can affect other people only through mediators that comply with the social rules. Their safety is guaranteed by the community so long as they use public mediators, as artifacts are designed to comply with the social rules, and hence are transparent at the granularity specified by the social rules. The responsibility flaw problem will be almost, if not complete, solved in principle in this framework, in the sense that the infrastructure is underwritten by the community responsible for the user.

It should be noted, however, we should admit that even the society of public mediators is inherently incomplete. First, it cannot be free from accidental malfunction. Even worse, the above framework deteriorates the moral in crisis and overdependence on artifacts problems. The moral in crisis problem will be escalated, as the surrogates will do anything on behalf of the user, and as a result, people will lose opportunities to think about morals and practice in the society. The overdependence on artifact problem will become desperately serious to the degree unrecoverable. It appears that we have to live with being assisted with artifacts (Nishida 2009b).

Occasionally, the mankind might encounter unprecedented disasters, and the computational framework might cease to function as an infrastructure. That is the time for the mankind to depend on itself. Under the circumstances, a reasonable goal might be to create a mutual dependency between empathy and technology: Using technology to help people cultivate empathy among people so empathy in the society may allow people to help each other to restore the infrastructure of the civilization, should they suffer from disasters and breakdowns that might be caused by incompleteness of technology. Substantial portion of technology should be devoted to help the mankind foster the sociality and moral for that time.

3 Empathy is the central issue

Empathy has been defined as "the ability to understand others' emotions and/or perspectives and, often, to resonate with others' emotional states," or as "an affective response that is identical, or very similar, to what the other person is feeling or might be expected to feel given the context: a response stemming from an understanding of another's emotional state or condition" (Eisenberg et al. 2010). Empathy can also be considered to be equivalent to conviviality that allows individuals to identify with each other thereby experiencing each other's feelings, thoughts, and attitudes and hence is deemed a central concept to design a community (Caire 2009).

Empathy is critical for people to understand and to help each other to restore the infrastructure should there be a technology breakdown. Empathy should also be established between humans and artifacts to bring about better symbiosis.

In order to see the issues in more detail, let us consider a situation illustrated in Fig. 1, where a person in the right (A) talks about a difficulty and her friend (B) expresses a concern, a typical scene of empathy. There may be a wide variety of possible causes, ranging from the physiological to the ethical. The response of a friend might have come out of a mere politeness without accompanying much
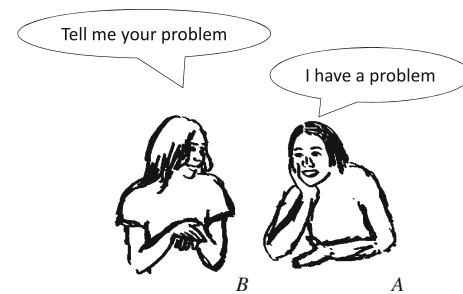

Fig. 1 Hypothesized mechanism of empathy
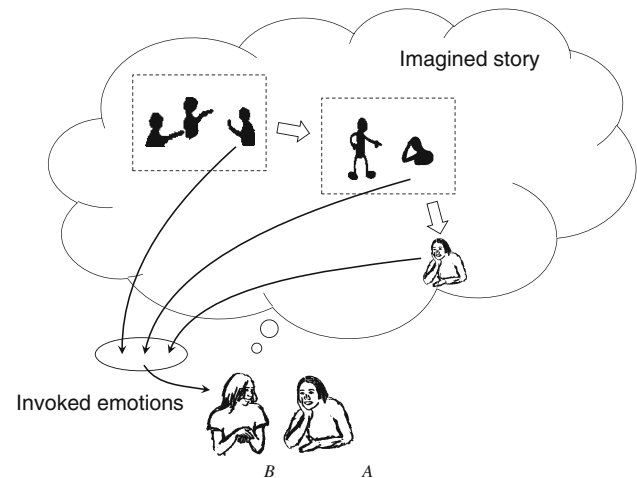
Fig. 2 Imagine what has happened from what is heard. *A* might have been discouraged and wanted to talk with *B*, for she was criticized not only by colleagues but also by her boss

thought. On the contrary, it might have come from concern at a much deeper level resulting from compassionate thoughts. For example, *B* may have imagined what has happened from what she has heard from *A* and felt in the way *A* might have felt (Fig. 2). In order to reach this level, one needs to sense and interpret social signals or subtle cues caused by a partner (Pentland 2008). According to theory of mind (ToM) (Baron-Cohen et al. 1985), people normally attribute mental states, such as desire or intention, to interpret and predict the behavior of other people. In general, reasoning about the mental process needs to be supported by the background knowledge. How can one acquire knowledge for it?

According to the theory hypothesis in cognitive development (Gopnik 2003), children develop their everyday knowledge of the world in the same way as adults develop scientific theories. A computational model for this hypothesis might be realized with a classic framework of symbolic knowledge-based abductive reasoning that will maintain one or more line of defeasible explanation for the observed behaviors of a partner, supported by the given knowledge base (Fig. 3). Unfortunately, however, this approach may run
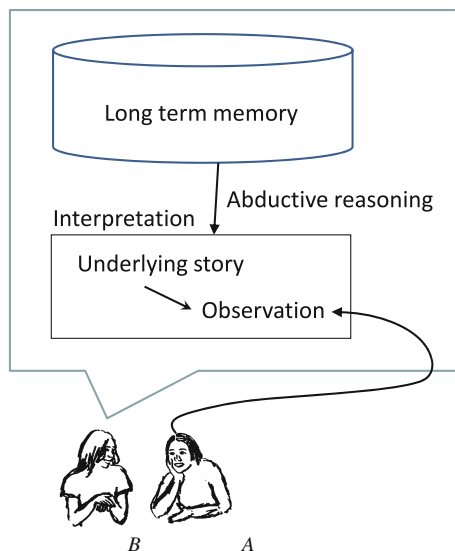
Fig. 3 Symbolic knowledge-based abductive reasoning

into a difficulty when little evidences are available for choosing among many possible interpretations.

The simulation theory suggests that human uses her/his own embodiment and mental model to simulate the behavior and mental process of the other people to understand them. For example, *B* may want to examine each step of the story she guessed by replacing the image of *A* by herself in the story, using a model about herself. *B*'s nervous system centered on limbic system will allow her to share emotional reactions similar to what *A* has experienced (Damasio 1994).

Recent findings of neuroscientists suggest that mirror neurons help us understand the emotions of other people by some form of inner imitation. Mirror neurons fire both when one performs an action and sees that action. They were first found in the ventral premotor cortex (area F5) of the macaque brain and then in the human brain (Rizzolatti and Sinigaglia 2008). Gallese et al. (2007) suggest that intentional attunement or embodied simulation enabled by the dynamics of our embodiment and neural system, including mirror neurons, might cause empathy. According to (Iacoboni 2008), mirror neurons help us understand the mental state of other people by making some form of inner imitation to pretend to be "in other people's shoes" (the mirror neuron hypothesis of empathy).

The computational model that reflects neuroscientists' arguments on empathy might be the one shown in Fig. 4. The architecture of an individual person's cognition may include an internal theater. In addition to a routine work of human information processing that takes information from receptors and produces motor commands, the input is sent to the internal theater where what happening is reproduced for, for example, reflection and planning. The communication partner's behavior is reproduced with the assistance
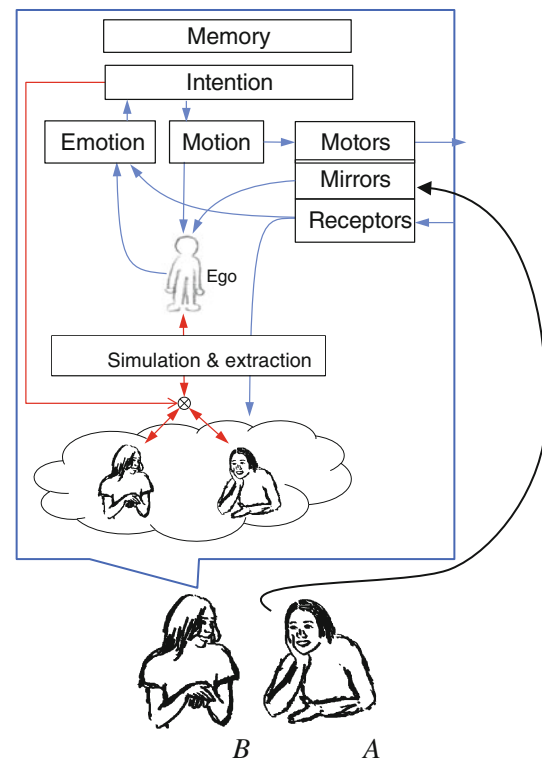
**Fig. 4** The computational model that reflects neuroscientists' arguments on empathy

of mirror units in the internal theater and interpreted by using the actor's own mechanism for generating emotional appraisals.

Unfortunately, the above framework does not retain all the advantages of symbolic knowledge-based abductive reasoning. In particular, it does not allow for contrasting one interpretation against another, which is often considered useful in figuring out the most plausible interpretation.

After all, the theory and simulation theory components need to be integrated into a coherent mechanism as shown in Fig. 5, where the two components play complementary roles to each other. The theory component will handle hypothesis making and reasoning at the cognitive level, while the simulation theory components carry out simulation with imaginary embodiment and emotions to evaluate hypotheses. The theory component guides and controls the simulation theory components, while the latter provides information about the appraisal of the given hypothesis based on the agent's own embodiment and physiology.

4 The blocking factors and sharing hypothesis

Even though people are gifted with a native mechanism for empathy, empathy may still not result due to blocking factors.
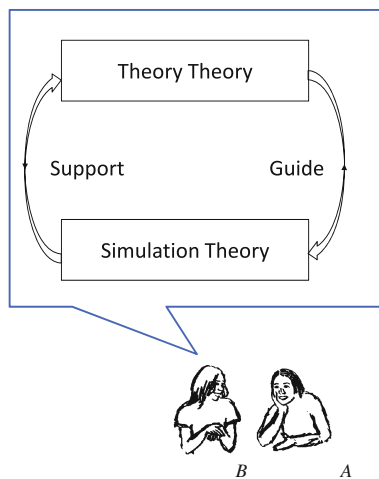
Fig. 5 An integrated model for empathy

First, the universe of discourse may have not been shared between the participants, preventing one from figuring out what has happened to the partner. Establishing the shared universe of discourse is not always trivial. When the participants are talking over the phone, for example, it is often difficult for them to reconstruct the shared universe of discourse, probably because the subject is not amenable to verbal description or because they cannot share the environment to which they may refer to communicate their thoughts. Even in face-to-face conversation where the environment of conversation is shared, the participants may have difficulty in interpreting what is said about objects or events due to the lack of shared background.

Second, empathy might be blocked due to the failure in sharing first-person view with other people. It may prevent one from perceiving the universe of discourse from different perspectives and resolve conflicts caused by the discrepancy of the perceived universe. For example, you might not be able to feel how small children experience with the world unless you look at the world at the same altitude as their eyes; you might not understand the difficulty of people who lost sight until you try to move around in a hazard with your eyes closed.

Third, inferior knowledge or skill level might not allow one to interpret the meaning in the same way with an expert. Although one may enjoy the play by a professional, she/he might not able to perceive or be aware of events that make sense to experts nor affect the world in the way experts do.

Fourth, empathy may be distracted by the differences in communication style, or the way intentions are encoded into social signals ranging from the way how verbal and nonverbal signals are coordinated make up an utterance, to the way the discourse is structured to satisfy the speaker's intention. Unlike behaviors whose meaning can be inferred on the anthropological or ethological grounds, the meaning

of rituals (Goffman 1967) cannot be understood without being taught by a member in the community because they have resulted from arbitrariness the community has chosen from multiple choices. Failure in sharing communication style and rituals will not allow one to exchange clearly defined messages with the communication partners, and predict how other agents may behave in a given communication environment, resulting in blocking empathy. Mishap might be observed in intercultural communication where social signals generated by the speaker are interpreted by the hearer in a different way, which might cause double bind and distress of the hearer.

Finally, empathy may not be induced if there is a discrepancy in the way the value is determined for events and objects. According to cognitive appraisal theory (Ortony et al. 1988), emotion is determined by evaluation of incoming events according to the value system. At higher levels of the mental process, the value system is used to make decisions. Empathy might not be obtained unless the observed behaviors do not follow the value system believed or at least approved by the observer.

It appears that the more technology removes those blocking factors and the more is shared among the participants, the more empathy is gained [the *sharing* hypothesis: (Nishida 2011)]. Information and communication technologies have a large potential for bringing about various kinds of sharing that have not existed before the information age. The internet and web technologies have brought about significant impacts on helping people share the universe of discourse, first-person view, knowledge and skills, the communication style and rituals, and the value system. The sharing hypothesis emphasizes the potentials that technology may realize in the future, rather than enumerating a list of blocking factors. Indeed, we have a long list of challenges. In the rest of this essay, I will highlight an immersive collaborative interaction environment for helping people share first-person view to bring about empathy and its application to building empathic agents.

5 An immersive collaborative interaction environment for scaffolding the first-person view

The first-person view allows for looking at the world from the perspective of other people. In order to comprehend their mental processes, one needs to share not only the first-person view like a life-log, but also a view surrounding them and the options available for them.

An immersive interaction environment as shown in Fig. 6 may partly satisfy the above mentioned requirements, for it permits one to experience a first-person view at a remote place to feel as if she/he were there. The remote cylindrical surface, called a *bubble*, provides the user with
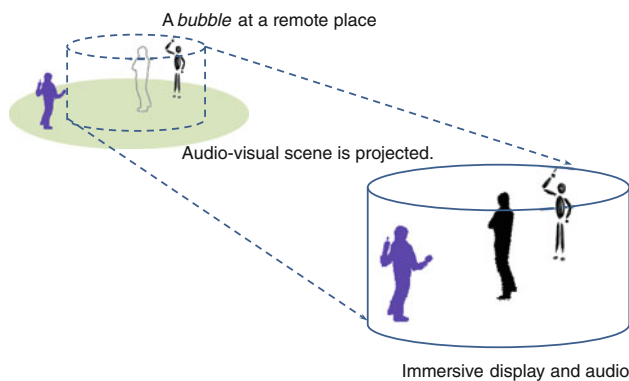
Fig. 6 Projecting a remote omnidirectional view to an immersive display
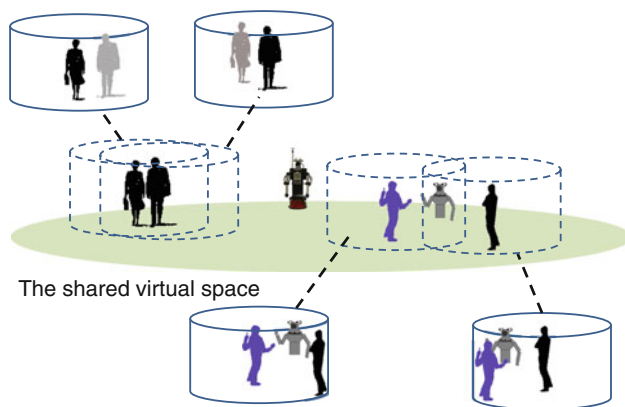
Fig. 7 Interacting bubbles in the shared virtual space

the audiovisual scene around her/him, allowing her/him to virtually occupy there to collect signals around her/him. One may attach an omnidirectional camera and microphones to a mobile robot that moves around the physical environment or virtually embed the bubble into a shared virtual space, as augmented virtuality (Milgram and Colquhoun 1999) shown in Fig. 7. It will permit one to create and share the universe of discourse beyond spatial and temporal constraints.

ICIE (Immersive Collaborative Interaction Environment) (Ohmoto et al. 2011) is an implementation of the above idea. The current version of ICIE employs eight 64-inch display panels arranged in a circle with about 2.5 m diameter. Eight surround speakers are used to reproduce the acoustic environment. The ICIE can be connected to a remote audiovisual capture (AVC) device so that the user in the immersive environment can feel as if she/he were standing at the place the AVC device is located.

The audiovisual scene at a remote place can be projected into the ICIE screen in either online or offline. In the online mode, scenes captured by one or more CCD camera or even omnidirectional camera possibly attached to a mobile

vehicle can be projected on the screen on the fly. It will produce a real-time immersive image as if the user in ICIE were moving around the remote place. A tele-presence environment using ICIE will help one dwell (Polanyi 1966), if not completely, in the mind of an agent who occupies the bubble and interact with her/his colleagues.

In the offline mode, we have developed a method for reconstructing a virtual space consisting of panoramic images for a given area of the real world, by combining multiple computer-vision techniques such as structure from motion, multiview stereo, and depth map (Mori et al. 2011). As the user virtually walks around the given space, a panoramic image for the location is computed almost in real time. The current version can automatically reconstruct from approximately 1,200 digital photographs, a 3D scene for a 20 m × 20 m space in 4 days (Fig. 8). It will help the user formulate an embodied image of the target place and familiarize her/himself there, which might help her/him share spatial cognition of the target place for empathic communication.

Recently, Masaharu Yano has prototyped a portable 3D multiparty conversation recording environment (Yano 2012). It integrates static recording of environment and 3D recording using multiple range sensors. It not only allows for re-constructing a 3D movie for multiparty conversation but also tracking a first-person view moving with a given actor (Fig. 9), which permits one to understand how other people may observe the world from a given perspective [virtualized reality (Kanade et al. 1995)]. Work is in progress on real-time mapping of the first-person view on the immersive display of ICIE.

ICIE can be used as an interface for a virtual space. We are building a system called *simulated crowd* (Thovuttikul et al. 2011), which is based on the idea of synthetic culture, that is, "role profiles for enacting dimensions of national culture" (Hofstede et al. 2002). Simulated crowd allows

Fig. 8 The user walks around a virtual space corresponding to a place of 20 m × 20 m in our university campus
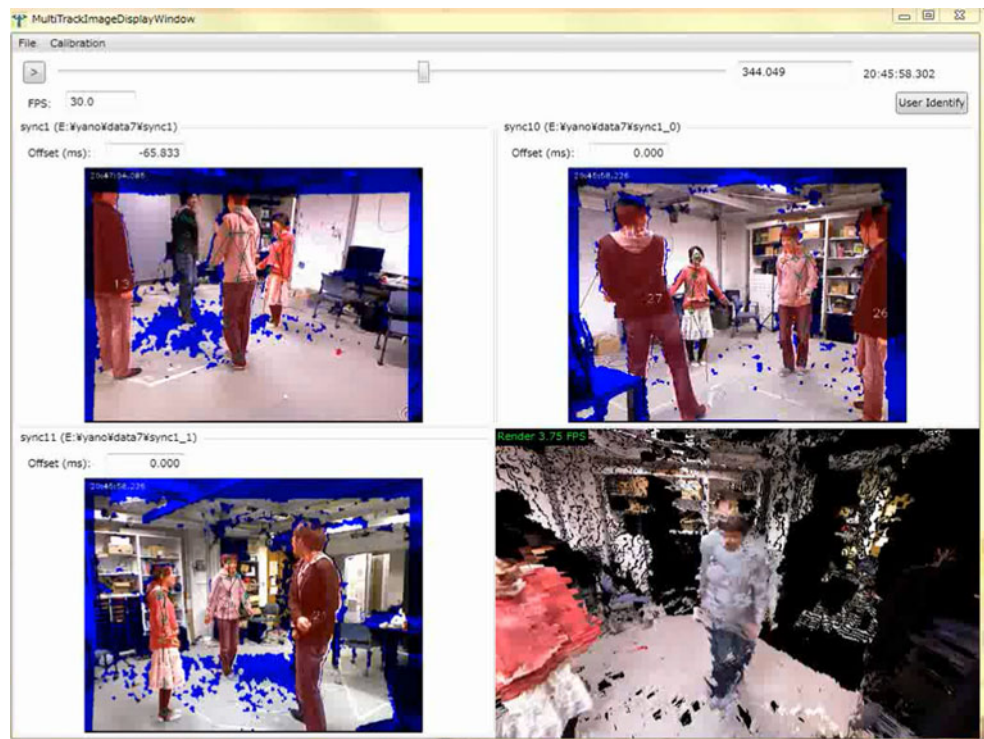
people to practice culture-specific nonverbal communication behaviors. A simulated crowd can be characterized as a possible instantiation of synthetic culture, which specifies an artificial environment habited by synthetic agents behaving according to a parameterized norm. Simulated crowd allows the user to virtually walk in a crowd of a given culture as a form of mental programs (Hofstede 2001) to experience culture-specific nonverbal signals by which people cooperate or negotiate with each other to avoid collision and achieve the respective goals (Lala and Nishida 2011c) (Fig. 10). We have found that the different settings of values for generic parameters may influence not only the physical characteristics of the place (such as average travel time or wait time in the space) (Lala and Nishida 2011b) but also cultural interpretations by the user

Fig. 10 The user interacting with a virtual crowd

(Lala and Nishida 2011a). The major future challenge involves intelligent tele-presence that can mediate the communication between the user and remote participants by semi-autonomously navigating the user's avatar and interpreting the situation at the remote conversational field for the user (e.g., language translation and annotating social cues).

In summary, ICIE extends the way people share the first-person view. It allows them to look at the world from other perspectives, contributing to the increase in opportunities of sharing and, hence, empathy.

6 Empathic agents

An *empathic agent* is defined as an autonomous agent that can create and maintain empathy with people. Empathy between humans and artifacts is a key to build human-artifact symbiosis when we are surrounded by an unlimited number of autonomous artifacts in a computationally mediated society. Once completed, empathic agents as a surrogate will not only efficiently and securely understand our intentions to provide with maximal service as our fellow surrogate but also motivate humans to understand the incompleteness of autonomous agents. It should be noted that empathic agents have not been introduced to directly resolve the moral in crisis or overdependence on artifacts problem, even though I believe that empathic agents may indirectly contribute to those issues.

Building an empathic agent has been deemed as one of the most challenging problems in AI, for it involves reproduction of a substantial portion of a mental process of humans. From the perspective of the sharing hypothesis, the developments in AI can be characterized as long-standing efforts on building autonomous agents that can share, if not comprehensive, meaning and intellectual functions with people.

SHRDLU (Winograd 1972) was an early natural language understanding system for a "block" world. The user can use English to tell SHRDLU how the block world should be changed. Given an input TTY-type utterance from the user, SHRDLU will figure out the meaning of the utterance by referring to linguistic knowledge concerning syntax, semantics, and discourse, and invoke a planning system in search for a plan for achieving the given goal. Thus, SHRDLU allowed the universe of discourse to be shared with the user, based on linguistic and task knowledge.

Another natural language dialogue system developed earlier than SHRDLU was Eliza (Weizenbaum 1966) Although Eliza did not understand the input in an ordinary sense, it invoked strong empathic reactions. The major reason was because Weizenbaum programmed Eliza in such a way that a standard TTY-based communication style was shared with the user.

Sharing communication style and rituals appears to induce empathy, as pointed out in the arguments for believable agents (Bates 1994) and the virtual theater (Hayes-Roth and Doyle 1998). Nonverbal expressions encompassing facial expressions, bodily expressions, and paralinguistic expressions are considered to be critical factors to play a critical role in demonstrating lifelikeness and personality. Nonverbal expressions are widely incorporated in embodied conversational agents (Cassell et al. 2000; Prendinger and Ishizuka 2004; Nishida and Nishida 2007).

Emotional expressions such as pleasure and pain might be universal in humans and animals, as pointed out by Darwin (1872), and hence deemed to be indispensable in the definition of empathy. Even a simple machinery may give a strong emotional influence to humans (Breitenberg 1984). Based on (Damasio 1994), Pickard suggested a computational architecture with primary and secondary emotions. The lower level processing is based on signal and pattern information processing, while the upper level employs reasoning and decision making to interface conceptual information representation with emotional behaviors (Pickard 1997).

An integrated system FAtiMA-PSI was proposed, which allows the user to experience with the first-person view of a specified character in interactive dramas (Aylett and Paiva 2012). FAtiMA-PSI includes a mechanism for modeling other agents and their relationship to the individual agent, which is able to build and update a record of the motivational state of other agents according to events perceived. Symbols, rituals, and appraisal are used to represent cultural aspects.

Unfortunately, those attempts have several severe limitations in establishing empathy with humans. First, conventional autonomous agents cannot become significantly smarter by being taught to meet new demands for sharing. Even though some adaptive learning mechanisms allow them to improve their performance, they do not contribute to acquiring intelligent/communication functions. In contrast, humans (after a certain age) are very efficient in extending a repertoire of their intellectual skills by simply being taught. Second, the conventional autonomous agents cannot rely on their own embodiment and physiology to interpret the internal state of the communication partner in novel situations. In contrast, humans can handle novel situations with the assistance of their embodiment and physiological mechanism.

In the next section, we will discuss the first issue, while we discuss the second issue as a future problem in Sect. 8.

7 Teaching by indwelling and learning by mimicking for sharing

According to the sharing hypothesis, an autonomous agent needs to effectively acquire what is required to be shared for empathy. Although the most effective transfer of knowledge among humans is teaching, autonomous agents are not easy to teach. What hinders teaching appears to be caused by failure in sharing first-person view between the human teacher and the agent learner, as the human teacher cannot successfully grasp how the agent learner curves the world and what it needs to learn from the teacher. In this section, I discuss how ICIE introduced in Sect. 5 can be integrated with a learning-by-mimicking engine to allow for effective agent learning from the human-assisted interaction, which we believe is effective in acquiring tacit knowledge (Polanyi 1966).

Our immersive wizard of Oz (WOZ) environment on ICIE allows human's expertise to be transferred to robots in the situated fashion. ICIE allows the human operator to feel as if she/he stayed inside a robot, so that she/he can choose the most plausible communication behavior under the circumstances. At the same time, the immersive WOZ environment captures the WOZ operator's behavior in real time using multiple infrared range sensors and maps it on the remote robot (Fig. 11). The sound on each side of the WOZ operator is gathered by microphones and sent to the remote robot to produce utterances (with a modulation, when necessary) (Ohashi et al. 2010).
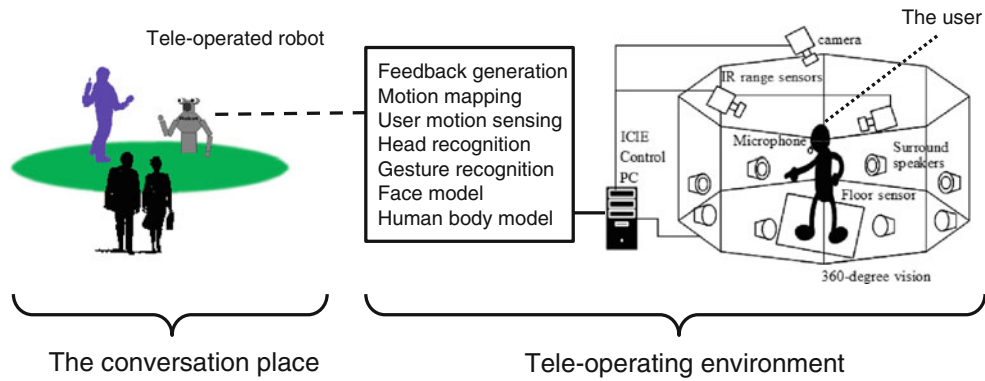
**Fig. 11** The immersive WOZ environment on ICIE

The behavioral model of the robot is inferred from the collected data through four stages of learning (Mohammad et al. 2009; Mohammad and Nishida 2010). At the discovery stage, the basic actions and commands will be discovered. At the association stage, a probabilistic model will be generated to specify the likelihood of the occurrence of observed actions as a result of observed commands. Granger causality is used to discover natural delay. At the controller generation stage, the behavioral model will be converted into an actual controller to allow the robotic agent to act in similar situations. Finally, the gestures and actions learned from multiple sessions will be combined into a single model at the accumulation stage (Fig. 12).

In order to handle mental processes at higher levels, we need to address how much one could help the user formulate a preference structure on an unfamiliar subject by repeating interviews consisting of presentation of possible choices and asking for the preference (Ohmoto et al. 2010). Incomplete verbal responses and measurement of body movement and physiological indices (SCR and LF/HF) are integrated to estimate the ordered set of features the user emphasized during each interview. We take into account that the user's preference structure may change as she/he changes emphasized feature. Preliminary experiments suggest that our method can track the user's changing emphasis and often result in user's satisfaction. A further experiment (Ohmoto et al. 2011) suggests that critical changes of features emphasized by the user can be detected by the combination of incomplete verbal reactions, body movements, and physiological indices.

It should be noted that I do not intend to claim that technologies presented in this section or those in Sect. 6

Fig. 12 A framework for learning by imitation (Mohammad et al. 2009)
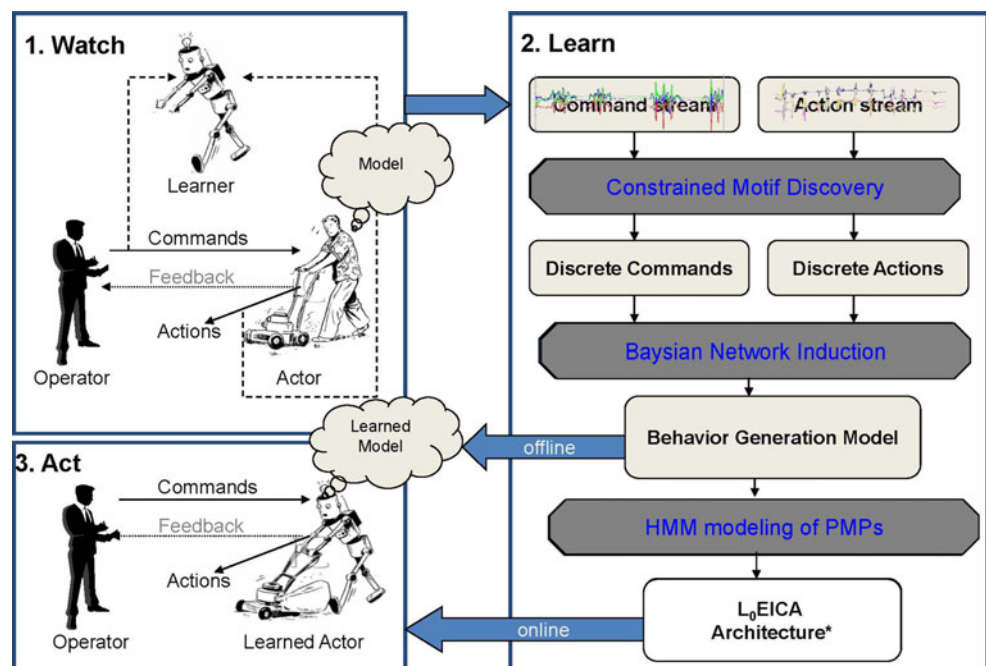
have proved advantages over the previous system. Instead, I have overviewed them in order to show how technology might contribute to cultivating emotions among people or realizing empathic agents.

8 Future perspectives

The most challenging issue in realizing empathic agents would be sharing embodiment and emotions. Although we know that certain techniques may increase believability or suspension of disbelief (Bates 1994) due to intentional stance (Dennett 1987) or the media equation (Reeves and Nass 1996), it appears that we need not only scratch the surface but also seek a way for elaborating the complex link between embodiment, emotion, and rationality.

A potential approach to resolving the discrepancy in embodiment between humans and artifacts appears to consist in simulation and metaphor. On the one hand, we might use physiological/mental simulations to reproduce the mental process at some levels of representation abstracted from the neuro-physiological levels. On the other hand, we might attempt to alleviate the discrepancy with the power of strong metaphor (Lakoff and Johnson 1980), so that shared metaphor should allow one concept to be represented by understanding of another concept, bridging over significant disparity in embodiment of humans and artifacts. We might depend on metaphor to project the embodiment and emotion of the human to the empathic agent and vice versa. Although the empathy between human and agent might be weaker than the empathy among people, the sharing hypothesis suggests that it would be much better than nothing.

In order to make a steady progress, we need to establish a methodology of evaluation by introducing an appropriate set of indices that can somehow measure how much empathy is cultivated as a result of a given technology. A full-scale evaluation shall follow once such indices have been figured out even partially.

9 Concluding remarks

In the former half of this essay, I have argued why technology will have to depend on empathy. First, I have pointed out that we will eventually live with super intelligence consisting of mediated autonomous agents as a result of techno-plosion brought about by info-plosion. Although a framework of computationally mediated autonomous agents appears to be an asymptotic goal of symbiosis between humans and artifacts, two problems will not be resolved: moral in crisis and overdependence on artifacts. In order to cope with this innate incompleteness

of super intelligence, we need to foster empathy in the mankind to create mutual dependency between empathy and technology.

In the latter half of this essay, I have discussed how technology may cultivate empathy among people. Following the discussion on empathy from computational points of view, I have introduced the sharing hypothesis, suggesting that empathy might be achieved by increasing what is shared. I have overviewed how an immersive collaborative interaction environment helps one experience situations from different perspectives.

Empathic agents are critical in realizing symbiosis between humans and artifacts. I have discussed empathic agents from the viewpoint of the sharing hypothesis. AI research can be seen as a history of unceasing efforts on realizing artifacts that can symbolically share with us perception, language, conceptualization, and knowledge, if not completely, by means of pattern recognition, natural language understanding and generation, knowledge representation and reasoning, and knowledge-based system.

A couple of challenges have been addressed to build empathic agents: making agents teachable and overcoming the discrepancy in embodiment and physiology. I have introduced the teaching-by-indwelling and learning-by-mimicking framework using an immersive collaborative interaction environment.

Introduction of a powerful mechanism of metaphor might be a key to overcome difference in embodiment and physiology. A potential future direction for empathic agents is to build a sophisticated human simulator encompassing embodiment and emotion and to embed an elaborate theory of metaphor for sharing between humans and artifacts.

References

Aylett R, Paiva A (2012) Computational modelling of culture and affect. Emot Rev (to appear)

Baron-Cohen S, Leslie AM, Frith U (1985) Does the autistic child have a 'theory of mind'? Cognition 21:37–46

Bates J (1994) The role of emotion in believable agents. Commun ACM 37(7):122–125

Breitenberg V (1984) Vehicles: experiments in synthetic psychology. MIT Press, Cambridge

Caire P (2009) Designing convivial digital cities: a social intelligence design approach. AI Soc 24(1):97–114

Cassell J, Sullivan J, Prevost S, Churchill E (eds) (2000) Embodied conversational agents. MIT Press, Cambridge

Cooley M (2007) From judgment to calculation. AI Soc 21(4):395–409

Damasio AR (1994) Descartes' error: emotion, reason, and the human brain. Gosset/Putnam Press, New York

Darwin C (1872) The expression of emotions in man and animals. Murray, London

Decker M (2008) Caregiving robots and ethical reflection: the perspective of interdisciplinary technology assessment. AI Soc 22(3):315–330

Dennett D (1987) The intentional stance. MIT Press, Cambridge

Eisenberg N, Eggum ND, Di Giunta L (2010) Empathy-related responding: associations with prosocial behavior, aggression, and intergroup relations. Soc Issues Policy Rev 4(1):143–180

Gallese V, Eagle MN, Migone P (2007) Intentional attunement: mirror neurons and the neural underpinnings of interpersonal relations. J Am Psychoanal Assoc 55:131–176

Goffman E (1967) Interaction ritual: essays face-to-face behavior. Aldine, Chicago

Gopnik A (2003) The theory as an alternative to the innateness hypothesis. In: Antony LM (ed) Chomsky and his critics. Blackwell, Oxford

Hayes-Roth B, Doyle P (1998) Animate characters. Auton Agents Multi-agent Syst 1(2):195–230

Hofstede G (2001) Culture's consequences: comparing values, behaviors, institutions and organizations across nations, 2nd edn. Sage, Thousand Oaks

Hofstede GJ, Pedersen P, Hofstede G (2002) Exploring culture, exercise, stories and synthetic cultures. Intercultural Press, Yarmouth

Iacoboni M (2008) Mirroring people—the new science of how we connect with others. Farrar, Straus and Giroux, New York

Kanade T, Narayanan PJ, Rander P (1995) Virtualized reality: concepts and early results. In: IEEE workshop on the representation of visual scenes, pp 69–76

Kant I (1788) Critique of practical reason (trans: Mary G). Cambridge University Press, Cambridge

Kitsuregawa M, Nishida T (2010) Special issue on information explosion. New Gener Comput 28(3):207–215

Lakoff G, Johnson M (1980) Metaphors we live by. University of Chicago Press, Chicago

Lala D, Nishida T (2011a) The interpretation of emotion based on gestural parameters and culture. CMVC 2011, Reykjavik, Iceland, 14 Sept 2011

Lala D, Nishida T (2011b) Towards a virtual environment for capturing behavior in cultural crowds. ICDIM 2011, pp 310–315

Lala D, Nishida T (2011c) VISIE: a spatially immersive interaction environment using real-time human measurement. GrC 2011, pp 363–368

Maurer H (2007) Some ideas on ICT as it influences the future. NEC Technology Forum, Tokyo

Milgram P, Colquhoun H (1999) A taxonomy of real and virtual world display integration. In: Tamura YOH (ed) Mixed reality: merging real and virtual worlds. Springer, Tokyo, pp 5–30

Mohammad YFO, Nishida T (2010) Learning interaction protocols using augmented bayesian networks applied to guided navigation. In: Proceedings of the IEEE/RSJ international conference on intelligent robots and systems (IROS 2010), pp 4119–4126

Mohammad YFO, Nishida T, Okada S (2009) Unsupervised simultaneous learning of gestures, actions and their associations for human-robot interaction. In: Proceedings of the IEEE/RSJ international conference on intelligent robots and systems (IROS 2009), pp 2537–2544

Mori S, Ohmoto Y, Nishida T (2011) Constructing immersive virtual space for HAI with photos. GrC 2011, pp 479–484

Nishida T (ed) (2007) Conversational informatics: an engineering approach. Wiley, London

Nishida T (2008) Towards robots with good will. In: Nagenborg M, Capurro R (eds) Ethics and robotics. IOS Press, Amsterdam

Nishida T (2009) Artifact-mediated society and social intelligence design. In: Bramer M (ed) Artificial intelligence: an international perspective. Springer, Berlin, pp 112–132

Nishida T (2011) Social intelligence design for intercultural communication. GrC 2011, Kaohsiung, Taiwan, 8–10 Nov 2011

Nishida T, Nishida R (2007) Socializing artifacts as a half mirror of the mind. AI Soc 21(4):549–566

Ohashi H, Okada S, Ohmoto Y, Nishida T (2010) A proposal of novel WOZ environment for realizing essence of communication in social robots. Social Intelligence Design, London, 13–16 Sept 2010

Ohmoto Y, Miyake T, Nishida T (2010) Estimating Preference structure through human–agent interaction, 3E2 − 2. In: Proceedings of the 25th annual conference of the Japanese society for artificial intelligence (in Japanese)

Ohmoto Y, Ohashi H, Lala D, Mori S, Sakamoto K, Kinoshita K, Nishida T (2011) ICIE: immersive environment for social interaction based on socio-spacial information. In: Conference on technologies and applications of artificial intelligence (TAAI 2011), Taoyuan, Taiwan, pp 119–125

Ortony A, Clore GL, Collins A (1988) The cognitive structure of emotions. Cambridge University Press, Cambridge

Pentland A (2008) Honest signals—how they shape our world. MIT Press, Cambridge

Perrow C (1984) Normal accidents—living with high-risk technologies. Princeton University Press, Princeton

Pickard WR (1997) Affective computing. MIT Press, Cambridge

Polanyi M (1966) The tacit dimension. Doubleday & Co., Garden City

Prendinger H, Ishizuka M (eds) (2004) Life-like characters—tools, affective functions and applications. Springer, Berlin

Rawls J (1999) A theory of justice. Oxford University Press, Oxford

Reeves B, Nass C (1996) The media equation. Cambridge University Press, Cambridge

Rizzolatti G, Sinigaglia C (2008) Mirrors in the brain: how our minds share actions, emotions, and experience? Oxford University Press, Oxford

Thovuttikul S, Lala D, Ohashi H, Okada S, Ohmoto Y, Nishida T (2011) Simulated crowd: towards a synthetic culture for engaging a learner in culture-dependent nonverbal interaction. 2nd workshop on eye gaze in intelligent human machine interaction at IUI 2011, CA, USA, 13 Feb 2011

Weizenbaum J (1966) ELIZA—a computer program for the study of natural language communication between man and machine. Commun ACM 9(1):36–45

Winograd T (1972) Understanding natural language. Academic Press, New York

Yano M (2012) Construction of 3-dimensional recording environments for multi-party conversation with RGB-depth sensors. Master thesis, Department of Intelligence Science and Technology, Kyoto University (in Japanese)