Consideration of Machine Consciousness in the Context of Mental Therapy from Psychological and Sociological Perspectives

Tatsuya Nomura*†
*Department of Media Informatics, Ryukoku University
1–5, Yokotani, Setaohe–cho, Otsu, Shiga 520–2194, Japan
nomura@rins.ryukoku.ac.jp
†ATR Intelligent Robotics and Communication Laboratories
2–2, Hikaridai, Seika–cho, Soraku–gun, Kyoto 619–0288, Japan
nomura@atr.jp

Koichi Takaishi‡
‡Department of Clinical Psychology, Kyoto Bunkyo University
80 Senzoku, Makishima–cho, Uji, Kyoto 611–0411, Japan
ko-takaishi@po.kbu.ac.jp

Tatsunori Hashido§
§Graduate School of Corporate Information, Hannan University
5–4–33 Amamihigashi, Matsubara, Osaka 580–8502, Japan
mc03013@hannan-u.ac.jp

Abstract
Realization of machine consciousness has a lot of scientific and engineering implications such as clarification of organization of human consciousness, implementation of real humanoid robots and virtual human agents, and so on. However, the most important is not how machine consciousness can be realized, but how people feel for robots and software agents when they recognize that the robots and agents have their own consciousness, and how the society is influenced by the result. This paper discusses these problems from psychological and sociological perspectives.

1 Introduction
Realization of machine consciousness has a lot of implications. Scientifically, organization of human consciousness will be clarified through the process of the realization. From engineering perspectives, it will lead to implementation of real humanoid robots and virtual human agents. Then, these robots and agents will be popularized in daily-life. When this popularization of machine consciousness is considered, however, the most important is not how machine consciousness can be realized, but how people feel for robots and software agents when they recognize that the robots and agents have their own consciousness, and how the society is influenced by the result.

These problems are critical when machine consciousness is applied to psychiatric fields such as mental therapy. In fact, current machines not having consciousness such as empathy have been starting to be applied to mental therapy (Shibata, 1999; Hashimoto, 2001; Dautenhahn and Billard, 2002; Fujino, 2003; Dautenhahn et al., 2002). Turkle (1995) reported that artificial agent programs for psychiatry have socially been allowed since 1990’s, while synchronized by standardization of diagnosis and treatment in psychiatry. Moreover, a lot of studies on robotic therapy have recently been done, in particular, in Japan (for example, Shibata (1999); Tashima et al. (1999); Hashimoto (2001)).

In fact, humans are affected even by these machines. Reeves and Nass (1996) showed that experiences of human through artificial media including computers are essentially equal to real experiences, by application of theories in social sciences on human interaction and experimental methods in psychology. In other words, humans unconsciously react to machines in the same way as to humans even if it is consciously recognized that those whom they interact are just machines. Some important results by Reeves and Nass (1996) are summarized as follows:

- Humans tend to politely react to machines sending polite messages, and prefer machines sending messages of praise to those sending critical messages.
Humans tend to interpret even image objects consisting of simple lines as ones having characters by using the same dimensions as those for humans (dominant–obedient and cooperative–non cooperative). In addition, dominant persons prefer machines displaying texts in dominant ways and obedient persons prefer machines displaying texts in obedient ways.

Humans tend to be affected by social roles assigned with machines such as professionals, teammates, and genders.

Humans tend to firstly feel good or evil emotions for information even from artificial media, and have a bias to negative information on attention and memory in the same way as the real world.

Important is that these reactions of human are unconsciously evoked. As a source of these phenomena, Reeves and Nass (1996) assume that they are a trace of evolution of mental mechanisms in the ancient wild environment.

Moreover, Turkle (1995) investigated minds of people on computers by interviewing with a lot of people in some countries from 1970’s to 1980’s. Some important results are summarized as follows:

- As mechanisms of computers became more complex, that is, they lost “transparency”, users gave up trying to understand computers based on the physical functions.

- Furthermore, as interactivity of computers became increasing, people began to understand these interactive and nontransparent objects by analogy of mental states of humans, in other words, regard them as objects having mental states, which is not either just physical objects or living ones.

- Researches on artificial intelligence and biologically inspired models such as neural networks have positively been affecting this trend. As one of the results, artificial agent programs for psychiatry have socially been allowed since 1990’s, while synchronized by standardization of diagnosis and treatment in psychiatry.

In addition, Turkle (1995) mentioned Eliza effect showing that humans tend to overestimate intelligent capability of computer programs.

These statements imply that even current machines can affect mental states of human by using actions based on their characters and social roles, regardless of positive or negative direction. Moreover, they imply that machine consciousness may really be introduced in psychiatric fields in future. However, it has sufficiently not been investigated what influences these machines have to clients in mental therapy. It has not been denied yet that machine consciousness in the therapeutic contexts have some evil effects to the clients dependent on the cultural situations where they live.

This paper discusses what meanings machine consciousness has in the context of mental therapy, that is, what conditions machine consciousness should satisfy when it is applied to mental therapy, whether machine consciousness can have positive effects when these conditions are satisfied, and what happens when it is introduced in the current social situations, from some perspectives of psychology and sociology.

2 Necessary Conditions of Machine Consciousness in Mental Therapy

This section considers conditions that machine consciousness should satisfy when it is applied as a substitution of psychotherapists, from methodological perspectives of clinical psychology.

2.1 Judgment of Timing

First, it is considered to be valid in all methods of psychotherapy that important is timing in treatment. It means that it is necessary to execute appropriate treatment in appropriate time in psychotherapy. In other words, timing in psychotherapy is a key determining whether the treatment succeeds or not.

However, it is just therapists that judge timing. Therapists individually judge it and take appropriate correspondence for clients. This individual judgment needs the therapists’ insight to see through conditions in which the clients stay. This judgment is dependent on perspectives of the therapists on the clients and changed by the therapists’ methods and experiences in psychotherapy. Thus, even if a client has the symptom same as another client, it may not be guaranteed that a treatment method effective for the latter is also effective for the former.

This judgment of timing in psychotherapy should also appropriately be executed by machine consciousness. In other words, machine consciousness should have its own experiences and construct its own insight...
2.2 Rapport

Second, the paradigm of clinical psychology implies that construction of a well relation between a therapist and client needs sympathy, warmth, and beliefs of them. This problem leads us to a concept of “Rapport”. Rapport is a state of relations that persons feel friendly and can confidentially talk with each other. It is necessary between clients and therapists in psychotherapy and can exist on confidential relationships between them. Mutual confidence in persons requires their mutual understanding.

Thus, machine consciousness should construct confidential relationships with humans based on its emotion of empathy.

2.3 Pragmatic Analysis and Irregularity

Third, human communication is based on pragmatic analysis for sentences, which deals with “indirect” or “deep” meanings of sentences, in contrast with semantic analysis which deals with “direct” or “surface” meaning. In other words, natural language processing in processes of communication requires not only surface processing but also understanding meanings contained in the background, according to the situations. If this process is executed by machine consciousness, much knowledge of humans’ social behaviors and mental states are necessary.

Moreover, daily life conversation between humans includes irregularity not following grammar and contexts. In other words, it has rich unpredictedness, which means that it is not predictable what utterances appear. This irregularity may also be a hint of treatment (for example, there is a method in family therapy where this irregularity is explicitly explored). If machine consciousness copes with this unpredictedness, a wide range of exception handling must be performed. It should also be necessary to prevent machine consciousness from executing inappropriate replies for humans’ unexpected behaviors. This type of processing is the most important in psychotherapy.

Thus, machine consciousness should execute pragmatic analysis and cope with irregularity based on much knowledge of humans’ social behaviors and mental states.

3 Software and Robotic Therapy under “Psychologism”

Even if machine consciousness having functions mentioned in the previous section are realized, it does not mean that it is useful in mental therapy. Then, this consideration needs psychological and sociological perspectives.

The cultural trend called “psychologism” implies possibility of evil effects of machine consciousness to clients in mental therapy. This word refers to a trend in modern society where psychiatric symptoms in individuals are internalized although they may be caused by social structures and cultural customs, and, as a result, the root social and cultural situations that need to be clarified are concealed.

In this section, we refer to sociological criticism for psychologism and consider implications from it.

3.1 Criticism for Psychologism from the Sociology of Emotions and Clinical Sociology

Mori (2000) focused on psychologism on discussing the extreme self-control of people in the modern society. His theory is based on the theory of feeling rules by Hochschild (1983) and the theory of McDonaldization of Society (rationalization) by Ritzer (1996), and is summarized as follows:

- In the modern society we are always forced to pay attention to our and others’ emotions in order not to hurt our emotions each other (cult of personality). Moreover, this cult of personality and psychologism has been complementing each other.

- Furthermore, psychologism and rationalization in the modern society has also been complementing each other, and as a result we are required to have a high degree of self-control for our emotions.

- Persons executing a high degree of emotion management cannot permit others’ deviation from feeling rules they observe even if it is only a little. This strict observance of feeling rules and difference of the rules between individuals cause disagreement in the modern society (e.g., increasing child abuse in Japan).

---

2This consideration is based on Nomura and Tejima (2002); Nomura (2003b)
In addition, Mori (2000) claimed based on analysis of increasing psychological manuals for self-helping that psychological knowledge strengthens the social trend of self-control for emotions. Psychologism has also been criticized in the research field of clinical sociology. Ozawa (2000) criticized the trend that people in the modern societies are dependent on counseling due to psychologism and the extreme emotion management.

3.2 Implications for Machine Consciousness

The above statements from the sociology of emotions and clinical sociology imply that people in the modern society are always required to execute emotion management and dependent on mental therapy for it. Moreover, rationalism as Ritzer (1996) pointed out may also encourage reduction of man power in mental therapy. In addition, Turkle (1995) implies that software agents and robots may be introduced in psychiatric fields. As a result, software and robotic therapy with machine consciousness including emotions such as sympathy may be encouraged.

On the other hand, these people are sensitive for others’ emotion management and there is difference of feeling rules between individuals. Thus, these people are also sensitive for emotional behaviors in the software and robots and there is possibility that the emotional behaviors of the systems are not suitable for feeling rules of the clients. Furthermore, it is not clear how people feel for empathy from machines when they appear as substitution of human therapists since the people regard the machines as objects having mental states which is not either just physical objects or living ones, while sensitive to the machine consciousness and being unconsciously reacting in interaction with it (Reeves and Nass, 1996). Thus machine consciousness may give the clients mental burden of emotion management in interaction between them and the systems, and influence therapeutic effects even if they are exactly implemented along theories on therapy.

4 Machine Consciousness as a Popular Product in Mental Therapy

The previous sections consider influence of machine consciousness in mental therapy at a level of individuals. This section considers possibility that machine consciousness is supplied as a popular product in mental therapy, from some sociological perspectives.

4.1 Perspectives from the Sociology of Health and Illness

There is another sociological research related to psychologism, called “the sociology of health and illness”. According to Nomura (2000), the sociology of health and illness is one of reflective sociological research actions that analyze and criticize discourses on health, dominant in the modern societies, and its theory is based on social constructionism (Lupton, 1994). The research subjects in the sociology of health and illness are relations between cultures and ways of using the concept of “health”, mutual interaction between medical staffs and clients, powers that the concept of “health” can have in the societies. As one of sociological researches of health and illness, Ukigaya (2000) analyzed the social situation on life-style related disease, focusing on diabetes. In her research it was clarified that advertisement of the concept of life-style related disease by the government extremely requires individuals’ accountability for health, relations between appearance of the illness and the social situations are concealed as a result, and clients of diabetes are socially and mentally pressed under requirement of self-accountability for their health.

Moreover, it was reported that some clients of diabetes develop their original interpretation of medical knowledge on the illness, and distort it. For examples, some clients are not perfectly ruled by medical indication, such as on meals and sports, but have their original meals and sports according to their bodily and mental states.

4.2 Another Perspective from Narrative Therapy

On the other hand, Asano (2001) pointed out that a new type of mental therapy called narrative therapy (McNamee and Gergen, 1992) has a factor to become popular in the modern societies. He claimed as follows:

- The action to talk narratives on selves is one of cultural practices popular in the modern society, that is, there are a lot of increasing people to want to talk narratives on themselves in USA, Europe, and Japan.

This consideration is based on Nomura and Tejima (2002); Nomura (2003b).
• The modern society has a characteristic to produce these people, and industries aiming at satisfying demand of these people like manuals for making narratives on selves, publishers, and so on, called “narrative industries”, have appeared. We should note that narrative therapy is just one of these industries.

• Narrative therapy functions by explicitly drawing things concealed in narratives which clients talk on themselves. However, the desire of people to talk narratives on themselves is also a desire to leave these concealed things concealed. If narrative theorists are not conscious for this fact, narrative therapy has a danger that it only repeats this desire of people.

4.3 Implications for Machine Consciousness

The statements by Ukigaya (2000) have some important implications. The concept of life–style related disease and psychologism have the common social power in the sense that both of them press people with some symptoms under self–control of individuals and conceals social situations related with sources of the illness. As a result of it, clients of mental therapy, in particular, software and robotic therapy may develop their original interpretation of psychological knowledge that is presupposition in implementation of these systems, and sometimes distort it. Moreover, the statements by Asano (2001) imply that reception of machine consciousness in the modern society and narrative industries satisfying desires of people to talk on themselves are combined with each other, and as a result machine consciousness appear as a product to help people to make narratives on themselves through interaction with people.

These facts imply problems in cases that therapeutic software and robots are supplied as popular products, not via medical organizations. People having their original interpretation of psychological knowledge with distortion may tend to prefer popular systems suitable for their interpretation of psychological knowledge to systems that are scientifically investigated and selected via medical markets.

However, interaction with machine consciousness may just repeat desire of people to talk on themselves while leaving concealed things concealed in their narratives, which should be drawn in narrative therapeutic conversation between clients and therapists. This situation can happen since narrative therapy does not mean a concrete therapeutic technique but just an attitude that therapists should have for clients, and thus it has rooms for clients’ original interpretation for it.

5 Double Bind Situations in Mental Therapy by Machine Consciousness

This section deals with the deeper relations between the cultural trends of mental health and the clients’ personal traits in mental therapy using software agents and robots, that is, the relations between psychologism and anxiety traits for computers and robots. We then suggest that clients of mental therapy using machine consciousness may be forced into a kind of double bind situation (Bateson, 1972).

5.1 Anxiety for Computers and Robots

The concept of computer anxiety means anxious emotions that prevent humans from using and learning computers in educational situations and daily life (Hirata, 1990; Raub, 1981). Anxiety can generally be classified into two categories: state and trait anxiety. Trait anxiety is a kind of personal characteristics that is stable in individuals. State anxiety can be changed depending on the situation and time, and computer anxiety is classified into this category. From the perspective of education, computer anxiety in individuals should and can be reduced by educationally appropriate programs, and several psychological scales for its measurement have been developed (Hirata, 1990; Raub, 1981).

On the other hand, from the perspective of mental therapy, computer anxiety can influence the therapeutic effect of software therapy using machine consciousness. If the client’s computer anxiety is high, it can prevent interaction with the therapeutic software agents even if the agents are designed based on theories of mental therapy. Of course, communication anxiety should be considered even in therapy with human therapists (Pribyl et al., 1998). However, it can be reduced during the therapy process by the therapist’s careful treatment. It is not clear whether computer anxiety can be reduced during the therapy processes by software agents. Thus, anxiety should either be reduced before therapy proceeds or another person should assist clients in interacting with the software agents during the therapy process.

Similar emotions of anxiety should also be considered in robotic therapy. Robotic therapy may be different from therapy using software agents in the

---

4This consideration is based on Nomura (2003a,b)
sense that robots have concrete bodies and can influence client's cognition. Thus, anxiety toward robots may be different from computer anxiety. However, it should be considered that anxiety may be caused by highly technological objects and communication with them. In this sense, anxiety toward robots is a complex emotion of computer anxiety and communication anxiety (Nomura and Kanda, 2003; Nomura et al., 2004).

5.2 Double Bind Theory

The Double Bind Theory, proposed as a source of schizophrenia from the viewpoint of social interactions in the 1950s (Bateson, 1972), argues that schizophrenia may result from not only impact on the mental level of individuals, such as trauma, but also inconsistency in human communication. The conditions for double bind are formalized as:

1. the existence of one victim (a child in many cases) and an assailant or some assailants (the mother in many cases),
2. the customization of cognition for double bind structures through repeated experience,
3. the first prohibition message with punishment,
4. the second prohibition message inconsistent with the first one at another level (inconsistent situations),
5. a third message that prohibits the victim from stepping out of the inconsistent situation (prohibition of the victim’s movement to a meta level of communication.)

It is pointed out that the double bind theory itself has largely not been developed in the theoretical sense since the 1970s (Ciompi, 1982), and there has not been enough empirical evidence showing that double bind situations are a source of schizophrenia (Koopmans, 1997). Even if the double bind situations are not a source of schizophrenia, however, the double bind theory has been applied in the clinical field as a basic concept of family system theory (Foley, 1986), and it is said that double bind situations frequently exist in daily life.

5.3 Implications for Machine Consciousness

The consideration in the previous sections imply that people in modern society are required to execute emotion management and are dependent on mental therapy for it. In addition, modern rationalism may also encourage a reduction of manpower in mental therapy, and, as a result, software and robotic therapy using machine consciousness may be encouraged. Then, people in modern society may be forced to face therapeutic software agents and robots by the social pressure of the self–control of their emotions and mental health, and rationalism, in particular, if mental therapy becomes a duty of members in organizations such as businesses and schools.

If these therapies are introduced without consideration of anxiety that individuals may experience in their interaction with computers and robots, however, they may cause double bind situations for these individuals, of which clients with high anxiety for computers and robots are victims. These clients are forced to face these systems by social pressure, but they cannot get sufficient therapeutic benefit due to their anxiety for the systems if their anxiety is not reduced by appropriate treatment, due to rationalism in the therapy process. Furthermore, social pressure prohibits them from stepping out of these situations because it signifies their rejection of accountability for their own mental health. In other words, this type of client cannot be treated with software or robotic therapy even if these software agents and robots are designed based on theories of mental therapy.

6 Summary

This paper considered problems of machine consciousness in the context of mental therapy from some perspectives of psychology and sociology.

As discussed above, even current machines without consciousness have a possibility of introduction in psychiatric fields while few sufficient investigation of their therapeutic effects. Even if they have consciousness needed for substitution of human counselors, it is not clear whether they have sufficient therapeutic effects, due to extreme sensitivity to emotions and social pressure of self–control for mental health caused by psychologism. Moreover, social pressure of self–control for mental health may encourage machine consciousness as a popular product of narrative industries in mental therapy, which just repeats desire of people to talk on themselves while leaving concealed things concealed in their narratives. Furthermore, introduction of machine consciousness in mental therapy at organizational levels may cause double–bind situations for clients having anxiety toward computers and robots, due to social pressure of self–control for mental health.

Of course, it should be discussed whether machine
consciousness can really have emotions such as empathy, in the sense that it has the same organization as humans. As far as humans regard software agents and robots as those having their own consciousness, however, its psychological and sociological influences should be considered, regardless that they can really have their own consciousness or not.

Finally, it should be noted that this paper does not aim at denying application of machine consciousness to psychiatric fields. There may be some disciplines familiar to software agents and robots in mental therapy. However, we should be careful of a naive idea that machine consciousness produces familiarity of humans with machines and realization of it leads to healing pains of clients in therapeutic contexts. As far as we consider applications of machine consciousness to therapeutic fields, we should pay our attention to influences of them in mental, social, and cultural levels.

References


the 8th International Conference on Artificial Life, pages 15–18., 2002.


