DYNAMICAL AND COMPLEX BEHAVIORS IN HUMAN-MACHINE CO-ADAPTIVE SYSTEMS

Tetsuo Sawaragi

Graduate School of Engineering, Kyoto University
Yoshida Honmachi, Sakyo, Kyoto 606-8501, Japan.

Abstract: Human and computer subsystems should be structured and designed to work in mutually cooperating ways guaranteeing a user's usability. For this purpose, progressive system redesigns are needed with respect to human computer interactions to increase system reliability and transparency by increasing human-system interactions and especially a human user's proactive participation, rather than by eliminating the human out of the loop. Such a *socially-centered view* on the human-machine system design regards a human and an automated agent as equivalent partners, and through their mixed-initiative interactions some novel relations of mutual dependency and reciprocity would emerge as well as flexible changes of role-taking are expected. To realize such a kind of new style of human-machine relationships, we develop a new idea called co-adaptive design principle, which means that both a human user and a machine should be able to adapt to the other through experiencing the interactions occurring between them. We applied this idea to an artifacts design of a robot tele-operation systems.

Keywords: Agents, dialogue, human-machine systems, human-centered design, human supervisory control, intefaces, interaction mechanism.

1. INTRODUCTION

The conventional division between human beings and machines should be modified in the context of thinking about *evolutionary* engineering processes. Human beings and the technologies including computers, communication devices, electronic networks, etc. should all be understood to be part of the system. Wherein, changes in the individual parts may take place through introducing alternate components, and all of these changes are part of the dynamics of the system. Sometimes such changes may be too complex for a designer to predict the behaviors emerging out of those.

Today's advanced automation might be indeed experts in solving/performing particular tasks, but have no means of relating to human users. Thomas Sheridan at MIT has called this “autistic automation” (Sheridan, 2002); “autism” represents those humans who seem to have lost their skill of becoming engaged, being embedded in a situation, a sense of belonging to the world and to their partners. Actually, such an aspect is becoming an origin of a new type of human errors caused by some mismatch between a human and machine autonomy (e.g., a well-known *automation-induced surprise* in aviation (Bainbridge, 1997). As a concept of human-centred automation (Billings, 1997) reveals, automation needs to behave "socially"; automation should learn a variety of powerful social rules which minimize interference and maximize group (i.e., human-automation) benefit and automation systems should be designed from the perspectives of "relations" and "processes" that may emerge out of the interactions between the automation and the human user.

In this paper, after surveying the problems incurred by the conventional technology-centered automation in a variety of fields, we put an emphasis on the fact that a concept of sociality is really needed to form the ideal relations of human-automation and to let them emerge out of intimate interactions. To realize such a kind of new style of human-machine relationships, we develop a new idea called *co-adaptive* design principle, which means that both a human user and a machine should be able to adapt to the other through experiencing the interactions occurring between them. We applied this idea to a variety of artifacts design of robot tele-operation and human-agent collaborative systems.

2. HUMAN-AUTOMATION DISCOORDINATION

One of the domains in which the most advanced automation is prevailing is an aviation domain, but the interactions among the pilots, air traffic controller and many automated devices may cause a new type of incidents and/or accidents initiated by a human error triggered by an usage of automation devices. The following is an overview of the actual accident of "near-miss," that was caused by the dis-coordination among a human air traffic controller and a pilot as well as an automated device of TCAS (i.e., warning device for aircraft collision avoidance).

In January 31, 2001, Japanese Commercial Airlines, aircraft A and B (Boeing 747-400D and a McDonnell Douglas DC-10) came within 10m of a collision in a near-miss incident in which the 747 crew ignored evasive action advice compounding suspected errors by air traffic control (ATC). On beginning a descent to 35,000 ft (10,675m) ordered by ATC, aircraft A's TCAS gave a serious 'RA' warning and the verbal order 'climb.' However the captain disregarded the advice and continued the descent. Meanwhile aircraft B's pilot was following descend advice from TCAS. But on seeing A's descending too, he started climbing back to 37,000ft. At some point during this maneuver both aircraft came within 10m as they crossed over. Air traffic controllers conducting training apparently gave confused instructions and repeatedly used the wrong flight numbers putting them on the same altitude and similar course.

The direct cause of the above accident is an air traffic controller's confused instructions, but this accident is revealing one aspect of limitation of current automation technologies. Human's judgment is "dynamic" in nature; his/her judgment is formed by a series of instructions by gathering cues to confirm his/her hypothesis under uncertainty. On the other hand, the judgment mechanism of the automation (such as a TCAS above) and its ways of conducting a human user is solely based upon a static status according to the predefined control logics, which may sometimes cause a conflict. This is illustrated in Fig.1.

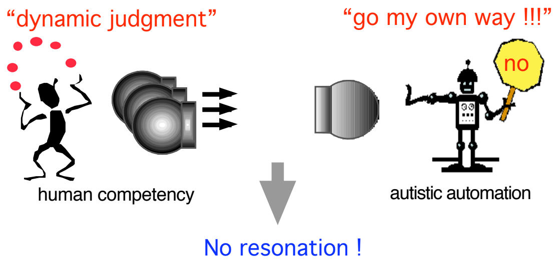

Figure 1. Autistic automation.

Here, we think that progressive system redesigns are needed with respect to human-automation interactions to increase system reliability and transparency by increasing interactions and especially a human user's proactive participation, rather than by eliminating the human out of the loop. Such a *socially-centered view* on the human-machine system design regards a human and an automated agent as equivalent partners, and through their mixed-initiative interactions some novel relations of mutual dependency and reciprocity would emerge as well as flexible changes of role-taking are expected. Our idea of *co-adaptive* design principle is illustrated in Figure 2.

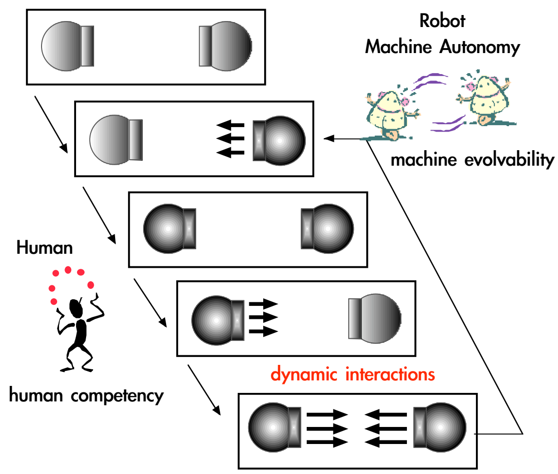

Figure 2. Co-adaptive system design principle.

3. HARNESSING: A NOVEL CONTROL AND DESIGN PRINCIPLE FOR COMPLEX SYSTEMS

In order to enrich the interactions between a human and machine autonomy and to let such a friendly and social relationships emerge through those interactions, we should abandon the conventional straightforward "control" doctrine and develop a novel principle for human-machine interactions. We think that a promising idea as an alternative to that is "harnessing", whose characteristics are described as follows.

(1) External input only gives direction for the path and its strength is kept as small as possible.

(2) Minimize the control input and let the system move by its own dynamics with reasonable resolution (i.e., not seeking for preciseness).

Machines to which this harnessing capability is embedded are assumed to generate a human-friendly mechanical behavior and as well as to present biological significance. For this purpose, machines should be evolvable though experiencing the interactions with a human, who is allowed to interact with a machine demonstrating a human competency.

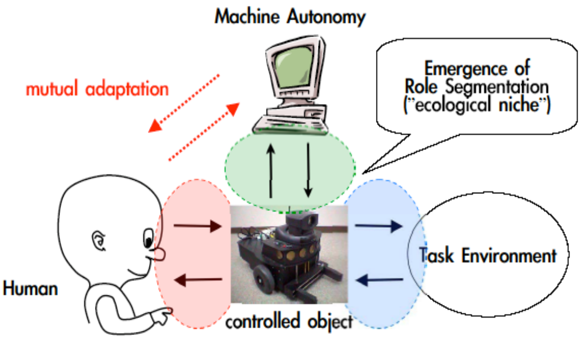

Figure 3. Tele-operated mobile robot.

4. SHARED AUTONOMY BETWEEN HUMAN AND MACHINE

As a testbed for constructing a human-machine collaborative system, we deal with a tele-operation system for a mobile robot as shown in Fig.3. We characterize this system as a shared autonomy system, meaning both machine autonomy and human autonomy must be shared. A main focus of a conventional simple tele-operation system has been attended to design an interface so that it could transfer an operator's control intentions and commands to a robot exactly as well as it could show a robot's behavior to a human as transparent as possible. Wherein, an ideal interface is the one that can establish a *morphological* mapping between a human task and a robot's task.

On the other hand, in a shared autonomy system both a human and a machine have their own autonomies, whose intentions are sometimes competitive and conflictive at least at the initial time. Through experiencing those conflicts and introspecting those competitions, both a machine and a human should be able to mutually adjust their judgments with each other and to find their own "niche" to perform collaboratively.

5. A GENERIC MODEL FOR CO-ADAPTATION BETWEEN TWO HETEROGENEOUS AUTONOMOUS AGENTS

As a generic model for such a co-adaptive process emerging in collaboration by two autonomous entities, we construct a model shown in Fig.4 as a pair of autonomous entities, each of which is a self-organizing system consisting of hierarchical structures simultaneously undergoing a variety of distinguishable activities. Different sets of variables and parameters are appropriate to a state space description pertaining to these activities taking place at the individual levels. Wherein, independence of descriptions of state spaces between the two entities is essential, and just a physical channel interconnecting them is shared. We do not assume that neither any "symbols" nor any "meanings" can be transferred on this channel, since symbols should be constructed and grounded by the autonomous entities by themselves in a self-enclosed way rather than by a system designer's ad hoc definition. We just assume that what can be shared between them should be restricted to information of an *object* level in terminologies of Pierce's classical idea of *semiosis*. Levels of *sign* and *interpret* must be left to the individual agents' efforts rather than determined by a system designer.

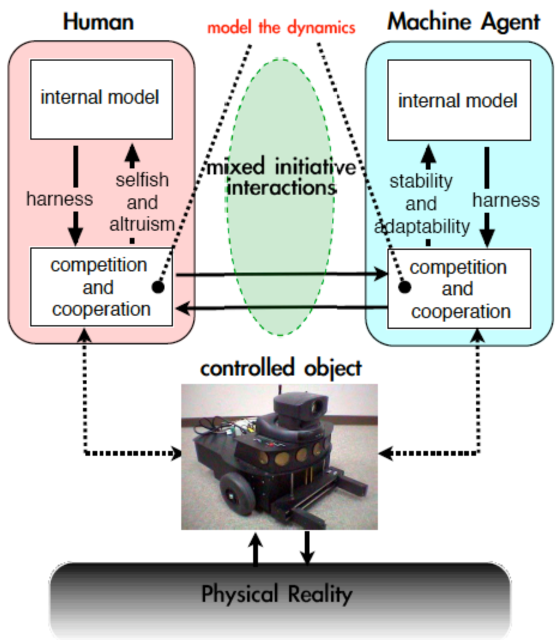

Figure 4. Human-robot Co-adaptation.

One of the key concepts of complex systems is a paradox that more than two competitive criteria may co-exist within a single entity as well as within an organization and/or team of them (Kaneko and Tsuda, 2000; Bar-Yam, 2002). This characteristic is making the behaviors generated both at a single agent level and at an organizational level dynamical and complex. That is, behavior formation cannot be implemented as a simple input-output function, and is quite different from the one by a classical stimulus-response model proposed in conventional behavioral psychology. A complex system behavior must be characterized by the key properties of "open" systems, where flows of matter, energy and information can occur across their boundaries, and this makes them undergo spontaneous transformations of structure and functionality within and among entities. Successive instabilities occur each time that existing structure and organization fail to withstand the impact of some new circumstance or behavior. When this occurs, the system re-structures and becomes a different system, subjected in its turn to the disturbances from its own non-average individual entities and situations. It is this interaction between successive systems and their own inner richness that provides the capacity for continuous adaptation and changes.

For realizing such a dynamics between two heterogeneous autonomous agents (i.e., a human and a machine agent), a hierarchical structure consisting of two layers is essential. Lower layer deals with the basic competitive and cooperative dynamics, and upper layer called an internal model maintains macroscopic status of its internal states evolving at lower layer from meta-level perspectives. In each of the autonomous entities, this basic architecture enables a reciprocal and bidirectional interactions. In the bottom-up direction some kind of order parameters constructed from the lower level is viewed as a representation of the internal model at a higher level, related with some macroscopic behavior ranging between two extremes of "autistic" and "social". In the top-down direction, on the other hand, only a few instructions and/or simple parametric commands are sent to a lower level intermittently, and then ongoing non-equilibrium statistical mechanics at the lower level may be affected indirectly and another equilibrium phase transitions may occur. In a word, the upper level takes a role of "harnessing" a dynamic behavior at the basic level by just adjusting a single parameter governing the dynamics at the lower.

In our framework, dynamics at lower level is implemented as non-constant-sum, nonnegotiable "Paradoxical" games in order to implement an "Ego" drifting between "selfishness" and "altruism" (Axelrod, 1997). This game is well-known as a Prisoner's Dilemma (PD) game and Chicken game (CG). In this model, each of the two players (agent 1 and 2) takes either of cooperation (C) or defeat (D) on the partner, thus their state is one of the four possible states of CC, CD, DC, or DD. For each state, payoffs that each of the players can get are defined as illustrated in Fig.5. In CC (both players take a cooperation), both of them can get a payoff of 1.0, but when either of the player takes defeat and the other takes cooperation, the payoff of a defeating player is ξ, while the payoff of a defeated one is - ξ. If both of them take defeat, the payoffs of the both players are reduced to -1.0 and this paradoxical outcome leads to "behavioral paralysis". If the payoff for DD is increased from -1.0 to -2ξ, then the state DC and CD become local equilibria since DD state is too expensive for the players to afford ("Chicken game").

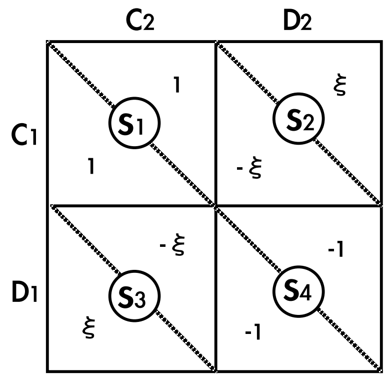

Figure 5. Payoff table for dynamics.

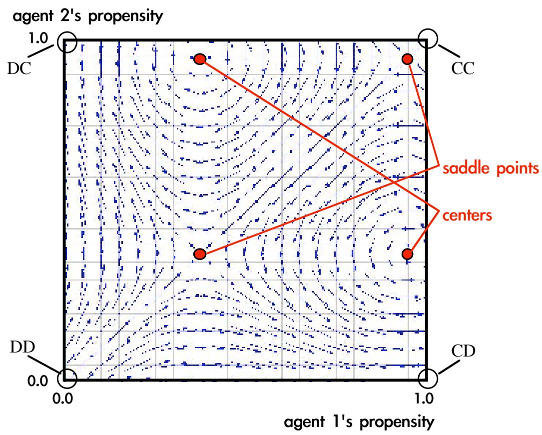

Figure 6. Evolution of transients of internal states along learning.

By assuming the games are played in iteration, learning takes place, which means that the time evolution of the propensities is governed by a system of two first order non-linear differential equations; the time derivatives of the propensities are proportional to the gradient of the expected payoffs with respect to that propensity. This is illustrated in Fig.6, in which a parameter ξ is set to a particular value. This figure shows; when two agents start from particular initial values of the propensities to keep taking cooperation, they get to be converged into one of the four states through iterating the games. Thus, this illustrates the internal dynamics of a *single* autonomous agent, in which competition and cooperation with its partner coexisting. This Markovian kinetics providing the basic dynamics at the lower level do not evolve under a fixed ξ (i.e., a game type), but the changeover between game types might be possible within each individual when the expected payoff with the previous game seems to be reaching local plateaus, and this changeover may be adjusted by changing a parameter ξ from the upper layer.

Then, how and when does this parameter of ξ should be adjusted from the upper level? It depends both on;

(1) Its own state: how consistent so far within itself (i.e., entropy of the state occupancy) that drives transitions of states at the higher level towards autistic attitude.

(2) Its partner's state: how consistent so far with its partner's (i.e., cross-correlation with the partner's behaviors) that drives transitions of states at the higher level towards social attitude.

Communication channel between the two autonomous agents is used for calculating the above cross-correlation. In other words, neither of any explicated intentions nor symbolic information is transmitted on it, but just cues that indirectly affect on both of the dynamics are transferred. Interpretation of those and how they are transformed into the adjustment of the parameter x are done in a self-enclosed way within the individual agents according to the above rules (1) and (2). The above modeling schemes are summarized in Fig.7 and Fig.8. The loops in Fig.7 show the dynamics of co-adaptation, each of which is modeled as consisting of two dynamics at upper and bottom layers in Fig.8.

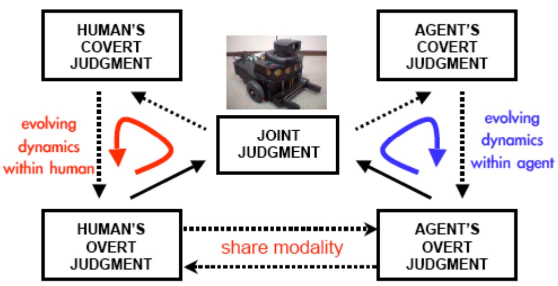

Figure 7. Evolving dynamics within each agent.

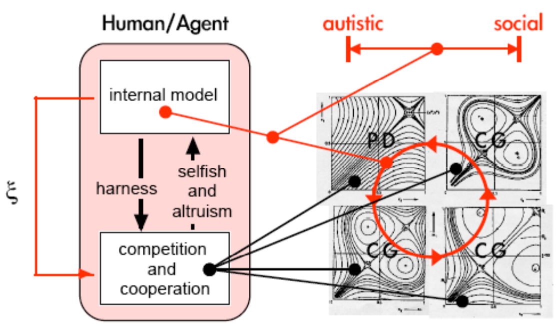

Figure 8. Two kinds of dynamics (slow and fast; upper and lower) within each agent enabling co-adaptation.

6. CONSTRUCTIVE APPROACH TO DESIGN AND CONTROL OF HUMAN-MACHINE CO-ADAPTIVE SYSTEMS

Based on the generic model for co-adaptation mentioned in the previous subsection, we are now taking a *constructive* approach to design and control of a variety of human-machine co-adaptive systems (Sawaragi, 2002). Constructing and simulating this model, we compare the model output with the reality. Based upon the above co-adaptation model, we are investigating into design of human-machine interface system for tele-operated mobile robot (Horiguchi, 2005).

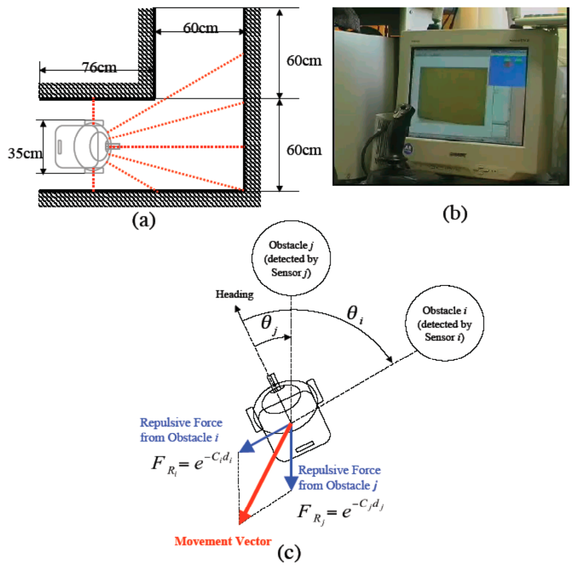

Figure 9. Testbed of tele-operated robot (a) task configuration (b) interface for a human (c) implemented machine agent's autonomy.

This study aims at a new design framework for combining and capitalizing on both advantages of the human and the mechanized automatic controls into their joint activity (i.e., shared control), wherein their well-coordinated collaboration is achieved through the interaction of dynamic and mutual shaping function allocations among them. Implementations of shared communicational modality between a human and a machine autonomy is realized by letting the intention of the robot autonomy transfer onto the joystick using the feedback force and by letting the operator's and the autonomy's input actions be mutually restricted through that joystick (Fig.9). This is called a Mixed-Initiative Interaction (MII). Machine autonomy is implemented by an adjustable potential field method composed of repulsive forces from obstacles that are caught by the range sensors mounted on the robot body. The joystick with the mechanism to generate the force-feedback effect is used to "embody" the model of the shared communicational modality. By letting decisions of the autonomous obstacle-avoidance behavior reflect on the joystick motions using the feedback force, the autonomy can also manipulate the joystick as well as the operator. Therefore, the operator's and the autonomy's input actions are mutually restricted through the joystick, since both of them can manipulate it and effect the other's judgment policies. Initiatives to control the robot can dynamically change as their inputs to the joystick become stronge.

We compared the collaborative performances with and without MII with respect to their steering operations through a series of experiments. The major findings obtained from these experiments are summarized as follows:

- Operators in the MII collaboration style could appropriately and consistently control their judgments in the joint operations with the robot autonomy, while in the No-MII condition their judgments might be disturbed.
- We investigated into the relative relation between human and mechanical judgment policies that are captured as linear models regressed from the experimental data. Operators under the No-MII condition evidently made more similar judgments with the autonomy's than the case under the MII condition. In other words, under the MII human and machine could find their *ecological niches*.

Then, we further embodied *proactive agency* into machine autonomy by letting it have an ability of making *probing* behaviors, by which the machine autonomy tries to get more information about the partner's covert decisions for its adaptation purposes. Iteration and accumulation of those interactions are expected to form some enduring processes toward their flexible or ever-changing collaboration with adequate mutual dependency and reciprocity. For this purpose, we developed an algorithm for machine autonomy, which consists of the following processes; 1) self-aware of perceptual discontinuity as conflicts with a human operator's judgment, 2) taking a proactive action to probe the partner's intention, 3) regard the partner's reaction as expression of its intention, and 4) adjust the own way of intervention. From the experiments, we could identify the following superior features of the above algorithm in contrast to a simple collaboration without any proactive agency embodied.

- Amounts of a human's operational fluctuation caused by his/her complementary operations to cancel the effects of machine autonomy's abrupt intervention are quite reduced.
- Human steering operations were performed more consistently in terms of the timing to start a turning operation because the robot gets to behave so that it can afford a human operator in an appropriate manner and timing.

Base upon the above experimental setting, we change design parameters for our human-machine system and observe the phenomena. What is important here is we are proposing a new design methodology based upon both the virtual events (i.e., using our co-adaptation model) and the actual events (i.e., using a testbed working in a reality). Recognizing the complexity of human-machine interactions as it is, we have to derive some universal design principles that can be mapped into the reality. We believe this constructive approach will be essential and rational way to explicate the covert truth existing in the complex system phenomena.

7. CONCLUSIONS

In this paper, we stressed an importance of automation's ability to form relations and to share a process with a human operator through intimate interactions. Proposing a novel design principle of co-adaptive systems, we analyzed a set of experimental results of human-machine collaboration and showed a way of constructive approach to the design of complex human-machine interactions.

ACKNOWLEDMENTS

The author appreciates my colleague Mr. Yukio Horiguchi, who is a former Ph.D. student supervised by the author. This work has not been accomplished without his efforts and contribution obtained as a part of his doctoral study. This work is attained under the 21st Century COE (Center of Excellence) Program of "Research and Education on Complex Functional Mechanical Systems" that is ongoing at Kyoto University.

REFERENCES

Axelrod, R. (1997). *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*, Princeton Univ Pr.

Bainbridge, L. (1997). The Change in Concepts Needed to Account for Human Behavior in Complex Dynamic Tasks, *Trans. of IEEE on SMC*, **27**-3, pp.351-359.

Bar-Yam, Y. (2003). *Dynamics of Complex Systems: Studies in Nonlinearity*, Perseus Books.

Billings, C.E. (1997). *Aviation Automation: The Search for a Human Centered Approach*, Lawrence Erlbaum Ass.

Horiguchi, Y. (2005). *Design of Co-Adaptive Interface System for Supporting Joint Task by Human and Machine Autonomies*, Doctoral Dissertation submitted to Dept. of Precision Eng., Graduate School of Eng., Kyoto University.

Kaneko, K. and Tsuda, I. (2000). *Chaotic Scenario of Complex Systems*, Springer-Verlag Berlin and Heidelberg GmbH & Co. K.

Sawaragi, T. (2002). Human-Automation Coordination Mediated by Reciprocal Sociality, *Preprints of IFAC 15th Triennial World Congress*, CD-ROM.

Sheridan, T. B. (2002). Humans and Automation: System Design and Research Issues, Wiley-Interscience.