Using self-aware agents to analyze public self-consciousness in the iterated prisoner's dilemma

Chung-Yuan Huang, Sheng-Wen Wang and Chuen-Tsai Sun
SIMULATION 2011 87: 600 originally published online 9 December 2010
DOI: 10.1177/0037549710391822

Simulation

Simulation: Transactions of the Society for Modeling and Simulation International
87(7) 600–615

DOI: 10.1177/0037549710391822
sim.sagepub.com

Using self-aware agents to analyze public self-consciousness in the iterated prisoner's dilemma

Chung-Yuan Huang[1,2], Sheng-Wen Wang[3] and Chuen-Tsai Sun[3]

Abstract

Self-aware individuals are more likely to consider whether their actions are appropriate in terms of public self-consciousness, and to use that information to execute behaviors that match external standards and/or expectations. The learning concepts through which individuals monitor themselves have generally been overlooked by artificial intelligence researchers. Here we report on our attempt to integrate a self-awareness mechanism into an agent's learning architecture. Specifically, we describe (a) our proposal for a self-aware agent model that includes an external learning mechanism and internal cognitive capacity with super-ego and ego characteristics; and (b) our application of a version of the iterated prisoner's dilemma representing conflicts between the public good and private interests to analyze the effects of self-awareness on an agent's individual performance and cooperative behavior. Our results indicate that self-aware agents who consider public self-consciousness utilize rational analysis in a manner that promotes cooperative behavior and supports faster societal movement toward stability. We found that a small number of self-aware agents are sufficient for improving social benefits and resolving problems associated with collective irrational behaviors.

Keywords

self-aware agents, public self-consciousness, cellular automata, small-world networks, tit-for-tat strategy, win-stay, lose-shift strategy

1. Introduction

Self-awareness is a psychological process through which individuals focus their attention on themselves.[1] According to self-awareness theory,[2] persons in high states of self-awareness contemplate the suitability of their own traits and behaviors, including aspects of their personalities, abilities, desires, needs, values, and other qualities.[3,4] Individuals use private and public self-consciousness to enact behaviors that meet standards that are internal, established by important others, or representative of societal and/or cultural values.[5] In other words, having an accurate sense of self-awareness supports a more complete understanding of the rightness/wrongness, goodness/badness, or superiority/inferiority of one's behavior according to social standards.[6] This is helpful for learning skills, managing personal interactions, and conducting one's affairs. Self-awareness can also support internal knowledge of emotions, motivations, interests, and desires so as to achieve self-identification and self-realization. In contrast, lack of self-awareness often leads to situations where personal behaviors are driven by strong emotions that are disadvantageous to one's well-being.[2] Further, a lack of understanding of others' emotions and ideas often leads to exaggerations of one's own merits and lack of restraint in social situations.

[1]Department of Computer Science and Information Engineering, Chang Gung University, Taiwan.
[2]Research Center for Emerging Viral Infections, Chang Gung University, Taiwan.
[3]Department of Computer Science, National Chiao Tung University, Taiwan.

Corresponding author:
Chung-Yuan Huang, Department of Computer Science and Information Engineering, Chang Gung University, 259 Wen Hwa 1st Road, Taoyuan 333, Taiwan
Email: gscott@mail.cgu.edu.tw

Since self-awareness plays an important role in human emotional development, some social psychologists are employing agent-based modeling and simulation techniques to explore the topic.[7–9] Descriptions of specific efforts can be found in the physics, biology, ecology, economics, management, computer science, and artificial intelligence (AI) literature.[10,11] Combined, these tools represent a novel approach to analyzing behavioral patterns at all levels of complexity in domains ranging from social development[12,13] to Internet commerce.[14,15] In the field of social psychology, which addresses complex human phenomena and societal processes, a growing number of researchers are using agent-based modeling and simulation tools to create multi-agent systems for exploring human behavioral patterns on a large scale.[16–18]

However, to date most forms of agent learning models have been restricted to what are essentially miniatures of external environments.[7] The emphasis has mostly been on mechanisms that connect stimulation signals and behavioral reactions.[19] Depending on the form of signals to which agents are exposed (primarily from physical environments or other agents), they use learning methods such as artificial neural networks, genetic algorithms, and fuzzy rule-based systems to adjust their knowledge bases or rule sets.[20,21] In some scenarios, agents are programmed to learn professional skills or to search for problem-solving strategies in response to pre-assigned demands or tasks.[22,23]

There are at least five advantages to using self-aware agents in artificial societies and social simulations.[11, 23,24] First, they are compatible with established AI learning frameworks in that they support extensions of those frameworks, increased agent learning performance, and faster arrival at optimal problem-solving strategies. Second, the introduction of a public self-consciousness concept lets agents take into account shared needs and feedback from other agents when building connecting mechanisms between behavioral reactions and stimuli – in other words, standards and values can be shared. Third, the presence of both private interests and public self-consciousness helps agents detect discrepancies between behavioral reactions and internal or external standards. They can therefore search for ways to decrease these discrepancies, increase learning performance, and satisfy internal or external standards. Fourth, an agent's ongoing and evolving discrepancies between the self and public self-consciousness can support researchers' attempts to understand external dynamics, resulting in timely revisions or adjustments in the agent's self-consciousness. Finally, self-aware agents support the construction of artificial societies and simulation models that more closely resemble those of real-world societies.

AI researchers have generally overlooked the learning concepts through which individuals enact self-awareness mechanisms.[8,9,23,24] We believe that integrating a self-awareness mechanism into an agent model will bring the behaviors and interaction patterns of agents into closer agreement with those of real people. Furthermore, self-aware agents will help agent-based social simulations more accurately reflect actual societal operations. Accordingly, we have developed a self-aware agent model aimed at improving learning capacity and performance. We will use the conflict between the public good and private interests to explore the influences of self-awareness on agent behaviors, degree of group cooperation, and larger societal interests.

2. Related theories and models

2.1. Self-awareness and social interactions

According to ethologist Desmond Morris,[25] self-concept represents the sum of self-beliefs. He suggests that self-concept affects all aspects of human personality formation, development, and change, and notes that in addition to their ability to think, humans have the power to self-analyze and to reflect upon and correct their ideas. Furthermore, humans are capable of understanding their feelings (i.e., to incorporate rationality into their feelings) and have the power of self-cognition, with which they can comprehend their positions in the world. Once established, human self-concept helps guide ideas, feelings, and behaviors[26] – for example, observing and making self-inferences via choices driven by internal motives, or doing what one *wants* instead of *has* to do. However, some researchers (e.g., Smith and Mackie[27]) disagree with claims that individuals can achieve self-understanding by observing their own behaviors. They argue that ways of forming self-images and ways of knowing others are very similar, but that humans have greater potential for deviance and for making mistakes in terms of self-perceptions.

In other words, it is easier to objectively observe one's neighbors than oneself – a power that can contribute to a cycle in which self-understanding helps one understand others and vice versa.[2] In addition, one's self-views are strongly influenced by the perceptions of others who are considered important or whose opinions are considered valid. Those perceptions help in the acknowledgment, interpretation, and correction of mistaken self-perceptions. This type of 'looking-glass self'[28] is very important for establishing appropriate attitudes and concepts, since it supports an understanding of how the world perspectives of others differ from our own. Self-recognition becomes blurred without this ability to use the eyes of others for self-reflection.[29,30]

2.2. Agent model

Agent architectures and knowledge (rules) are major factors affecting problem-solving speed and solution quality (performance). The two most commonly used intelligent agent-based models emphasize knowledge acquisition and adaptation.[7–9,19] The first is rule-based, with domain experts determining knowledge and rules required for agent problem-solving, which are integrated into the model in question so that agents can infer appropriate behaviors based on information from the external environment. The second is learning-based, with agents solving problems and collecting information from a combination of past experiences and present needs to construct new strategies or adapt old ones, and to transform strategies into knowledge bases and rule sets. Although rule-based agent architectures are easy to implement, they require large knowledge and experience data sets for use by agents, raising issues regarding revising, updating, customizing, and learning flexibility.

Current intelligent agent models combine these architecture types, resulting in adaptive models with built-in knowledge bases or rule sets that follow three steps: sensing, planning, and acting.[7] In Figure 1 we use the extended classifier system (XCS)[31] to illustrate the interaction process between sense-plan-act agents and their environments. Interactions between XCS (which is associated with genetic algorithms and reinforcement learning[22]) and a problem environment occur as follows: at discrete time t, an agent senses the current state s_t and compares it with individual rule or classifier conditions within an agent's internal knowledge base [P]. All rule sets that fulfill state s_t are stored in a temporary buffer [M]. Methods such as cost estimation and experience accumulation are used to select the most suitable rules in terms of fitness values, through which actions a_t are executed. Optimum learning is achieved through a system of rewards, punishments, and trial-and-error. Although scholars have proposed various learning algorithms for use with core intelligent agent architectures (e.g., artificial neural networks, genetic algorithms, Q-learning), sense-plan-act still serves as the foundation for most agent models.[7–9]

Stuart Russell[7] used a compilation of agent architectures to establish a general learning-based agent model. As shown in Figure 2, his model contains four components in addition to detectors and effectors that interact with the external environment: a learning component, a performance component that monitors the external environment and determines responses, an assessment component that evaluates learning performance, and a problem generator for determining optimal directions for additional learning. In addition to containing a detailed architecture for interactions between an agent's learning mechanism and external environment, Russell's model clearly addresses all critical components and connections in agent model design. We used his work as our self-aware agent model foundation.

2.3. Iterated prisoner's dilemma

Over the past decade, ecologists, anthropologists, political scientists, sociologists, economists, and computer scientists have used agent-based simulations with the non-zero-sum prisoner's dilemma to explore conflicts between the public good and private interests.[12,13] Researchers have also used the same combination to study the reciprocal effects, emergence phenomena, and evolutionary dynamics of rational strategies consisting of mutually beneficial cooperation and selfish defection.[32,33]

The payoff matrix for the prisoner's dilemma (PD) is shown in Table 1. $R = 3$ represents the reward for mutual cooperation, $T = 5$ one party's temptation to defect, $S = 0$ the 'sucker's payoff,' and $P = 1$ the punishment for mutual defection. Two key conditions for generating a PD are $T > R > P > S$ and $2 \times R > T + S$.[12] The first condition guarantees that two rational agents will simultaneously betray each other after understanding that $T > R$ and $P > S$, and therefore follow the second best choice: mutual defection (P, P). According to the second condition, prisoners cannot escape such a predicament by taking turns betraying each other – in other words, benefits for mutual betrayal are not as good as for mutual cooperation. In addition to these conditions, five PD criteria are thought to prevent all forms of cheating: (a) agents cannot communicate or negotiate with each other, (b) agents cannot threaten or make promises to each other, (c) agents cannot determine their opponents' future behaviors, (d) agents do not have the power to terminate play at any time, and (e) agents cannot change each other's payoff values. Accordingly, each agent can only rely on past behaviors to formulate strategies that optimize long-term benefits.

In each PD round, both individuals are free to choose cooperation or defection, with defection always holding greater potential for benefits. However, if both parties choose defection, their total compensation will be less than that generated if both choose cooperation. There are many real life scenarios in which participants encounter each other more than once. When those individuals recognize each other and remember past choices, then the prisoner's dilemma

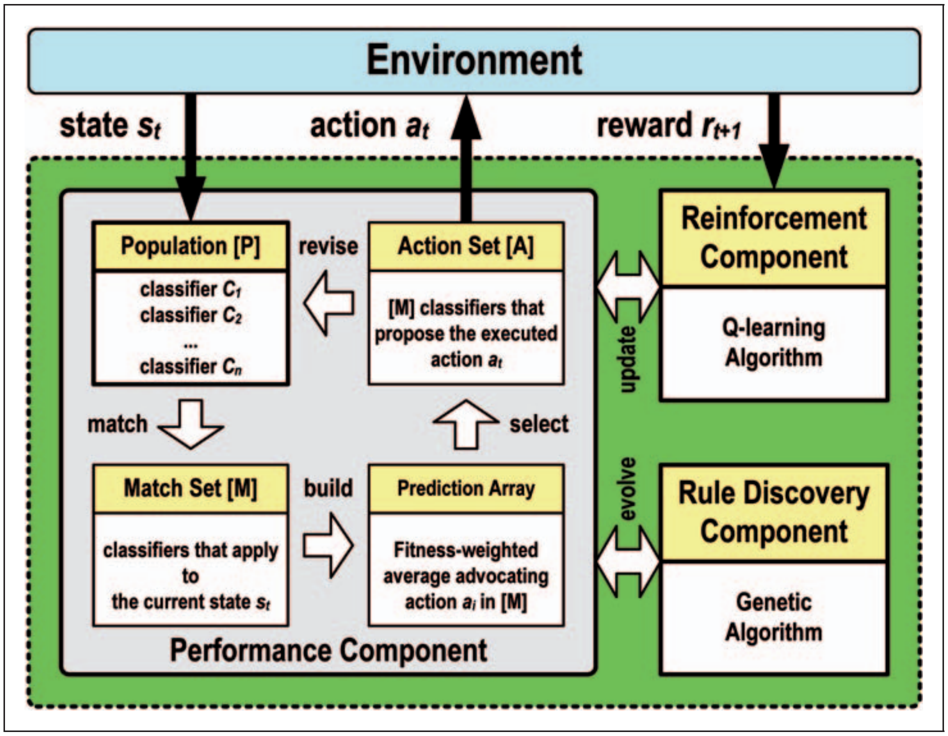

Figure 1. Extended classifier system (XCS) architecture, adapted from Huang and Sun.[22] XCS is a problem-independent and adaptive learning-based agent model with four components: finite classifier, performance, reinforcement, and rule discovery. Stored classifiers control the system via a horizontal competition mechanism, and perform tasks by means of vertical cooperation. The performance component governs interactions with the target problem. The input interface is used to transmit the state of the target problem to the performance component, and to determine dominant classifiers according to an exploration/exploitation criterion. All actions advocated by dominant classifiers are executed and receive feedback via the output interface. The reinforcement (credit assignment) component uses an algorithm similar to Q-learning to update the reinforcement parameters of classifiers that advocate the output action. The rule discovery component uses a genetic algorithm to search for better or more general classifiers, as well as to discard incorrect or more specific classifiers.

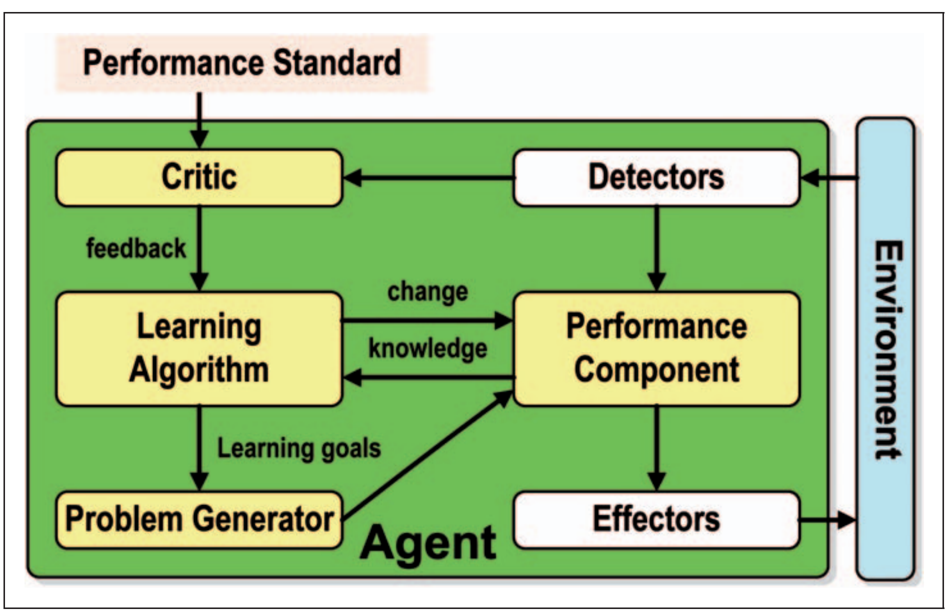

Figure 2. Russell's general learning-based agent model, adapted from Russell and Norvig.[7]

Table 1. Payoff matrix for the prisoner's dilemma, adapted from Axelrod.[13] Individual A's payoffs are in blue, B's in red

		B	
		Cooperation (C)	Defection (D)
A	Cooperation (C)	R = 3, R = 3	T = 5, S = 0
	Defection (D)	S = 0, T = 5	P = 1, P = 1

Table 2. Memory-1 deterministic strategies

No.	Strategy $(s_{cc}, s_{cd}, s_{dc}, s_{dd})$	Note
S_0	(C, C, C, C)	all-cooperation
S_1	(C, C, C, D)	
S_2	(C, C, D, C)	
S_3	(C, C, D, D)	
S_4	(C, D, C, C)	
S_5	(C, D, C, D)	tit-for-tat
S_6	(C, D, D, C)	win-stay, lose-shift
S_7	(C, D, D, D)	
S_8	(D, C, C, C)	
S_9	(D, C, C, D)	
S_{10}	(D, C, D, C)	
S_{11}	(D, C, D, D)	
S_{12}	(D, D, C, C)	
S_{13}	(D, D, C, D)	
S_{14}	(D, D, D, C)	
S_{15}	(D, D, D, D)	all-defection

becomes an iterated prisoner's dilemma (IPD). Interest in long-term survival encourages cooperation, since individual private interests do not generate the same benefits as group-oriented public concerns. In other words, both public good and individual fitness are best supported by long-term collaboration.

In addition to all-cooperation (ALL-C) and all-defection (ALL-D), two of the most analyzed IPD strategies are Nowak and Sigmund's 'win-stay, lose-shift'[34] and Rapport's 'tit-for-tat'.[12] The first is based on Pavlov's theory regarding the maintenance of one's current cooperation or defection strategy until benefits fall below a threshold value, at which point the opposite strategy is followed. In the second, agents always cooperate in the first round, and imitate their opponents' behaviors thereafter. According to Axelrod,[12] the four main characteristics of the tit-for-tat strategy are (a) friendliness, meaning that one does not defect before an opponent defects; (b) vengefulness, so that once an opponent defects, it is possible to immediately 'get even'; (c) tolerance, in that one stops defecting as soon as an opponent stops; and (d) transparency, making it easy for opponents to predict the results of a tit-for-tat strategy. By analyzing the evolutionary dynamics of various strategies and their predominance at different times during simulations, IPD researchers have found that reciprocal cooperation is an advantageous strategy that satisfies the condition of evolutionary stability.

As shown in Table 2, many IPD agents adopt a memory-1 deterministic strategy.[35] In other words, they use short-term memory to store the strategies and behaviors of opponents in the preceding round. There are only four possible combinations of behaviors for two agents: both cooperate (expressed as CC), the first cooperates and the other defects (CD), the first defects and the other cooperates (DC), and both defect (DD). This memory-1 deterministic strategy can be expressed as $(s_{cc}, s_{cd}, s_{dc}, s_{dd})$. For example, if an agent's memory of the preceding round is CC, then the agent will choose S_{cc} when responding to an opponent. Since each response is limited to either cooperation or defection, the memory-1 deterministic strategy has a total of 16 (2^4) possible moves, including $S_0 = (C, C, C, C)$, $S_1 = (C, C, C, D)$, $\ldots$, $S_{15} = (D, D, D, C)$, $S_{15} = (D, D, D, D)$. Among these, $S_0 = (C, C, C, C)$ is known as the 'yes-man' (ALL-C) strategy, $S_5 = (C, D, C, D)$ the tit-for-tat strategy, $S_6 = (C, D, D, C)$ the win-stay, lose-shift strategy, and $S_{15} = (D, D, D, D)$ the 'scoundrel' (ALL-D) strategy.

3. Self-aware Agent Model

Sigmund Freud described human personality as consisting of three parts: id, ego, and super-ego.[36,37] The id contains a person's subconscious thoughts, including primitive desires originating from intuitive impulses. The purpose of the id is to fulfill biological needs (including sexual desires) and to avoid pain. The super-ego (also referred to as the moral self or self-consciousness) is a personality monitor; it operates according to a 'perfection principle' that adheres to behavioral guidelines. A list of social rules and regulations that rely on the super-ego would include propriety, justice, honesty, shame, loyalty, filial piety, benevolence, love, trust, righteousness, harmony, and peace. The ego (the personality executor, or 'rational self') is the part of the personality structure that comes into contact with external environments. When a person's id and super-ego come into conflict, the ego uses a reality principle to make adjustments through which the needs of the id are satisfied, while super-ego rules and regulations are adhered to based on the current context. The ego makes use of multiple defense mechanisms to balance discrepancies between internal needs and external realities. For example, suppose someone finds a wallet containing $1,000. The id views this as a windfall, but the super-ego says the money should be returned to its owner. As an ego-driven compromise, the finder may decide to search for the person who lost the money for three days, and keep it if unsuccessful.

Adopting Freud's theory of personality to learning-based simulation agents presents many challenges. Current approaches emphasize continual agent adaptation to achieve optimal performance. In these cases, only the rational self (ego) is involved in helping agents to quickly react to environmental changes and achieve their goals. Although that is sufficient for survival, relying on ego only means that agents cannot act to achieve the greatest overall benefit for a community or society. In contrast, agents who only have egos and ids can achieve certain individual-centered tasks, but their self-interested behaviors and strategies will ultimately be rejected by other agents. Agents who possess both egos and super-egos can understand community or societal expectations and resolve public good/private interest conflicts; accordingly, cooperation among agents emerges more quickly so that societal benefits are increased. In a later stage of our research we incorporated the super-ego into our learning-based agent model, thus injecting a sense of self-awareness in support of agent efforts to understand social expectations.

Building on Russell's learning-based agent model, we developed a self-aware agent model consisting of three components: performance, learning, and cognition. As shown in Figure 3, the cognition component uses past experiences to add information to a knowledge structure in order to assist agents with understanding, explaining, and predicting self-behavior. The learning and cognition components, which are coordinated to improve performance, entail external environment (opponent strategies) and internal cognition features. According to our proposal, agents become capable of self-awareness through the addition of various schemas that improve efficiency via learning and cognition coordination, thereby moving closer to a model of human intelligence.

Since an agent's super-ego has the potential for increased social benefits and faster collaborative behavior,[38,39] we attempted to add this personality component to our agent self-awareness mechanism. This personality component can be analyzed as a mix of reputation and social expectation strategies. The idea of reputation, which has been extensively studied in the context of multi-agent systems,[40–42] can be used to judge opponent strategies for the purpose of making adjustments. Agent reputation is also considered useful for constructing self-awareness.

Reputation implementation consists of four steps:

Step 1: Assume that agent A_i has n opponents so that $O_i = (o_{i,0}, o_{i,1}, o_{i,2}, \ldots, o_{i,n-1})_g$ during generation g; $o_{i,j}$ represents agent A_i's j^{th} opponent ($j \in \{0, 1, 2, \ldots, n-1\}$).

Step 2: At the end of generation g we arrive at the numerical sequence $C_i = (c_{0,i}, c_{1,i}, c_{2,i}, \ldots, c_{n-1,i})_g$, with $c_{j,i}$ representing the number of times that opponent $o_{i,j}$ makes a cooperative move during a PD interaction with agent A_i in generation g.

Step 3: Calculate the average value (avg_i) and standard deviation (std_i) of the numerical sequence C_i.

Step 4: Assign the reputation value $r_{i,j}$ of agent A_i's opponent $o_{i,j}$ as $\lceil (c_{j,i} - avg_i)/std_i \rceil$.

Agents become aware of their fitness values when reputation mechanisms are introduced; at a certain point agents can determine their reputations based on judgments made by other group members. (The degree of social fitness of an agent has been widely discussed in the general multi-agent system literature (see, for example, Mead[28]). In the prisoner's dilemma, an agent's score can be used to indicate social fitness. However,

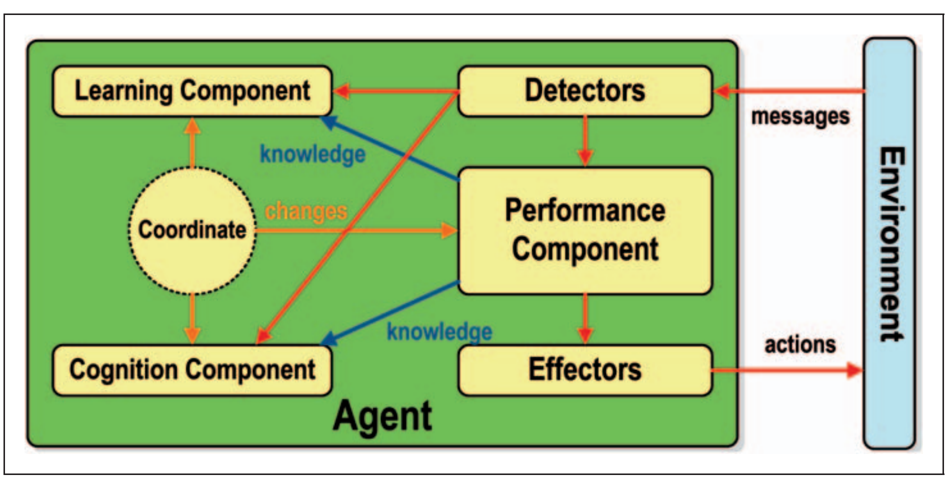

Figure 3. Self-aware agent model.

we believe this measure is insufficient, and that self-aware agents must respond to input from both the ego (social fitness) and super-ego (agent reputation).) As shown in Figure 4, fitness values and reputations can be divided into high, medium, and low categories (nine cross-categories in all). Using the combination of high fitness value and low reputation as an example, this agent type is likely to use an always-defect strategy in order to improve performance. However, an always-defect strategy does not improve the public good. The addition of a super-ego will likely trigger a self-adjustment, resulting in a strategy that simultaneously fulfills societal expectations and helps the agent achieve a higher fitness value. At the end of generation g, agent A_i can follow six additional steps to attain both a strong relative fitness value and reputation:

Step 1: Assume that af_i is agent A_i's fitness value.

Step 2: Agent A_i asks all O_i opponents about their evaluations of her reputation $r_{j,i}$. In the numerical sequence $AR_i = (r_{0,i}, r_{1,i}, r_{2,i}, \ldots, r_{n-1,i})_g$, $r_{j,i}$ represents opponent $o_{i,j}$'s evaluation of agent A_i's reputation. This information is used to calculate the average value ar_i of numerical sequence AR_i; the resulting ar_i represents agent A_i's average reputation in group O_i.

Step 3: Agent A_i collects reputation information for opponent O_i from all other opponents, and uses it to create the numerical sequence $OR_i = (r_{0,1}, \ldots, r_{0,n-1}, r_{1,0}, r_{1,2} \ldots, r_{1,n-1}, \ldots, r_{n-1,0}, \ldots, r_{n-1,n-2})_g$ with length $n \times (n-1)$. $r_{j,k}$ represents opponent $o_{i,j}$'s evaluation of opponent $o_{i,k}$'s reputation. Also as part of this step, the average value of numerical sequence OR_i (OR_avg_i) and standard deviation OR_std_i are calculated.

Step 4: Agent A_i collects fitness value data for all opponents O_i and creates the numerical sequence $OF_i = (f_0, f_1, f_2, \ldots, f_{n-1})_g$ with length $n \times (n-1)$. f_j represents the fitness value of opponent $o_{i,j}$. Also as part of this step, the average value of the numerical sequence OF_i (OF_avg_i) and standard deviation (OF_std_i) are calculated.

Step 5: If agent A_i's fitness value af_i is less than ($OF\_avg_i - OF\_std_i$), then $A_i.Fitness = LOW$. If the fitness value is greater than ($OF\_avg_i + OF\_std_i$), then $A_i.Fitness = HIGH$. If the fitness value is between ($OF\_avg_i - OF\_std_i$) and ($OF\_avg_i + OF\_std_i$), then $A_i.Fitness = MIDDLE$.

Step 6: If agent A_i's reputation ar_i is less than ($OR\_avg_i - OR\_std_i$), then $A_i.Reputation = LOW$. If the reputation is greater than ($OR\_avg_i + OR\_std_i$), then $A_i.Reputation = HIGH$. If the reputation is between ($OR\_avg_i - OR\_std_i$) and ($OR\_avg_i + OR\_std_i$), then $A_i.Reputation = MIDDLE$.

4. Simulations and Experiment Results

Complex systems such as biological ecosystems, human societies, and Internet commerce generally consist of large numbers of interacting and coordinating entities. When a small number of entities regularly betray, take advantage of, and/or appropriate others' resources in order to enhance their own performance, their self-interested behaviors will eventually damage the system's overall interests. As stated above, self-aware agents with egos and super-egos at the center of their operations are more likely to take into account their

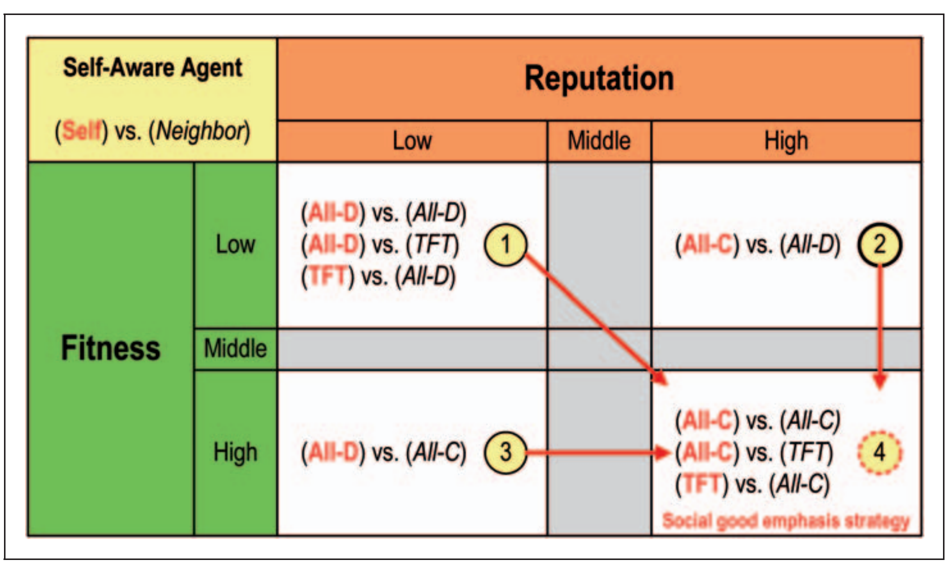

Figure 4. Agent fitness score and reputation index table.

own fitness values and reputations, and therefore search for appropriate strategies to boost community or societal performance while still supporting their private interests. We designed a simulation experiment with an IPD scenario to test the effectiveness of adding these personality characteristics.

Our simulation model includes two layers in which agent interactions take place: a multi-agent system upper layer, and a lower layer consisting of either cellular automata or a Watts and Strogatz (WS)[43] small-world network (Figure 5). Cellular automata are two-dimensional $W \times H$ regular networks with high degrees of local clustering and separation, while WS small-world networks have high degrees of local clustering and low degrees of separation. In order to compare IPD scenario results using these two network model types, we stipulated that node (agent) and edge (contact) numbers must be identical for the two network types. Every node (strategic agent) has on average eight adjacent nodes with which it comes into regular contact (e.g., family members, neighbors, fellow commuters, classmates, etc.). In cellular automata, each cell is equal to one node and edges are distributed according to a Moore neighborhood pattern. WS small-world network construction is identical to that of the cellular automata for the first stage. Once all nodes have been established and connected to their eight adjacent nodes, edge rewiring takes place according to a predetermined probability ρ.

Our IPD simulation experiment consisted of seven steps:

Step 1: Set environmental parameters and initial values for evolutionary computations (Table 3). Environmental parameters include total number of time steps for each simulation experiment, strategy and color mapping table, agent memory capacity, and human contact and interactive social network values (i.e., numbers of nodes and edges, neighbor patterns, and small-world network rewiring probability). Evolutionary computation parameters include total number of agents/nodes, crossover rate, mutation rate, and total number of generations/time steps.

Step 2: Use experimental requirements to choose and construct the most appropriate network – cellular automata or two-dimensional WS small-world.

Step 3: Set time step t to 0.

Step 4: Based on network connection patterns, have the nodes at the ends of links A_i and A_j execute q IPD rounds. Using the payoff matrix shown in Table 1, calculate scores for A_i (as_i) and A_j (as_j), both ranging from $q \times S$ to $q \times T$ (i.e., from 0 to $5q$). Use these scores as fitness values for agents A_i (af_i) and A_j (af_j), with $af_i \leftarrow as_i$ and $af_j \leftarrow as_j$.

Step 5: Agents calculate their relative fitness ($A_i.Fitness$) and reputation ($A_i.Reputation$) values, which represent the evaluations of all A_i opponents.

Step 6: Each agent determines whether or not to make strategy adjustments. The current strategy is considered inappropriate when $A_i.Fitness = LOW$ or $A_i.Reputation = LOW$. Agents that need to adjust their strategies use evolutionary computation crossover operations to combine their original strategies with strategies used by opponents with high fitness values. They also use evolutionary mutation operations to randomly change their strategies.

Step 7: Increase each time step value by 1. If $t < Time\_Step\_Limit$, then return to Step 4; otherwise, terminate the experiment.

A screen shot of our simulation system and user interface is shown as Figure 6. On the right is a list of parameters and their default values, on the left is the

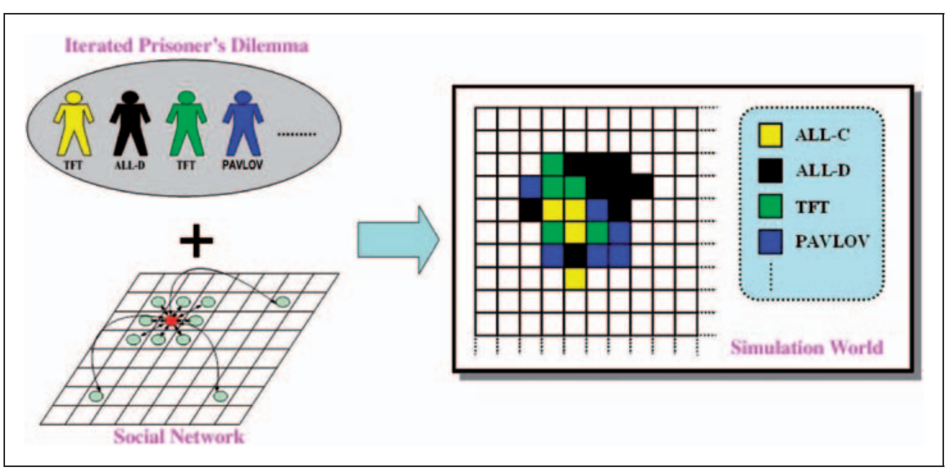

Figure 5. Our proposed simulation model.

Table 3. Experimental parameters

Parameter	Default value	Description
TIME_STEP_LIMIT	100	Total number of generations for each simulation experiment
q	100	Total number of interactions between an agent and its opponent
Network Type	$\times$	If *NetworkType* = CA, a 2D cellular automata is built; if *NetworkType* = WS-SWN, a 2D Watts and Strogatz small-world network is built
W	50	Width of 2D cellular automaton
H	50	Height of 2D cellular automaton
N	2,500	Total number of nodes (agents). Default value = $W \times H$
E	10,000	Total number of edges. Default value = $(50 \times 50 \times 8)$ / 2
Neighborhood	Moore	Von Neumann/Moore neighborhood pattern, with periodic boundary conditions
ρ	0.01	Specific parameter for a 2D WS-SWN. Generating such a network begins with 2D cellular automata with periodic boundary conditions. Each link is randomly rewired to a new node according to a rewiring probability
c	1	Agent memory capacity
P_c	0.7	Crossover rate of the genetic algorithm used in this work
P_m	0.01	Mutation rate of the genetic algorithm used in this work

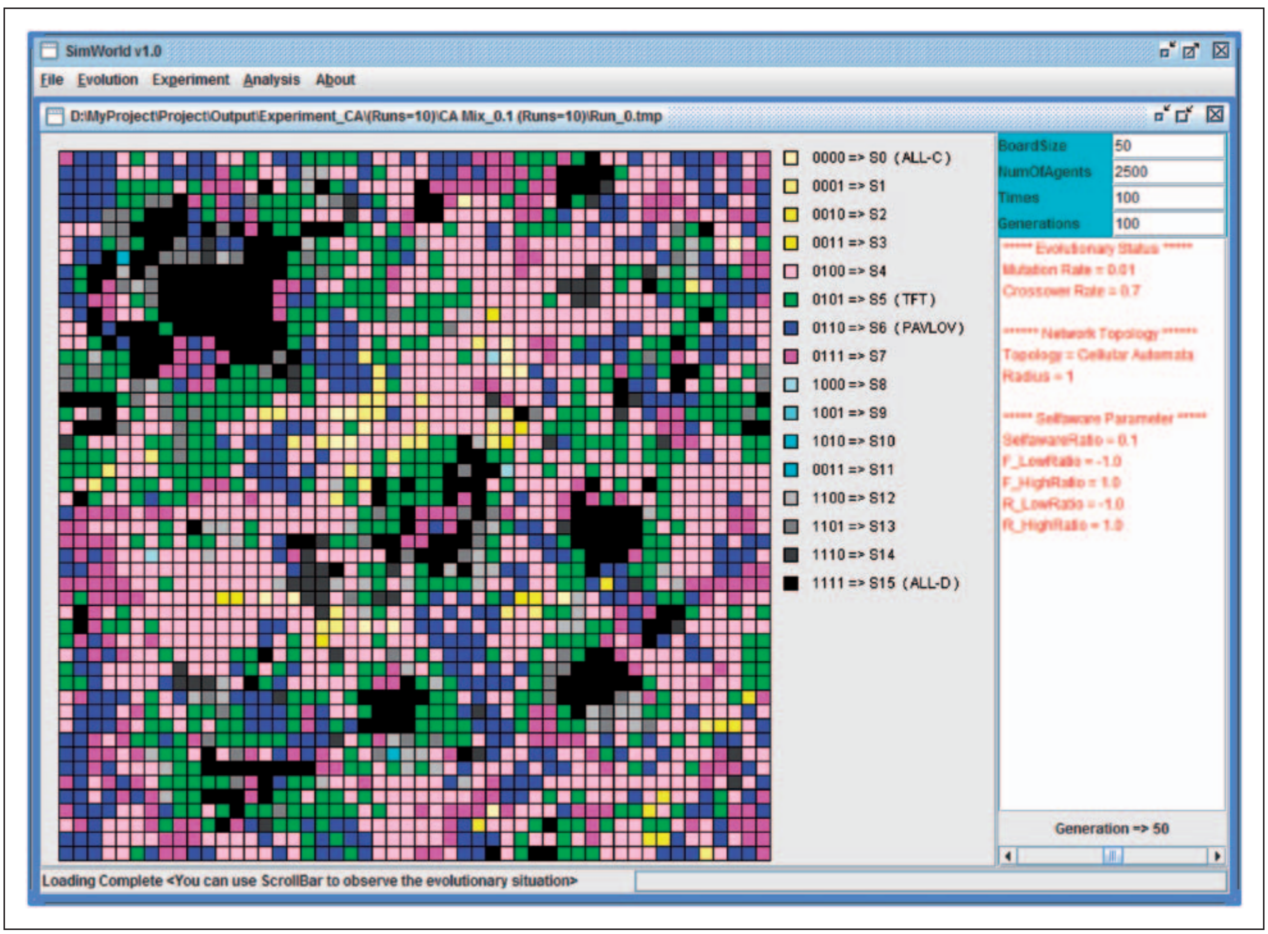

Figure 6. User interface for our iterated prisoner's dilemma simulation system.

underlying social network. Each cell (node) represents one strategic agent. To the right of the network is a list of sixteen memory-1 deterministic strategies and their corresponding colors. In a later section we will discuss four of the most commonly used strategies: ALL-C, TFT, PAVLOV, and ALL-D.

4.1. IPD simulation experiments

First, we spent time observing the complex behaviors of non-self-aware agents in IPD scenarios using cellular automata (Figure 7). The simulation process can be divided into five stages, based on the evolutionary dynamics and spatial distribution of agent-adopted strategies. During the first stage (generations 0 to 3), our proposed simulation model uses random numbers to determine initial strategies adopted by individual agents, resulting in those strategies being evenly distributed throughout the cellular automata (Figure 7(a)).

During the second stage (generations 4 to 10), agents tend to move toward the 'scoundrel strategy' (always defecting) due to the temptations of maximizing their

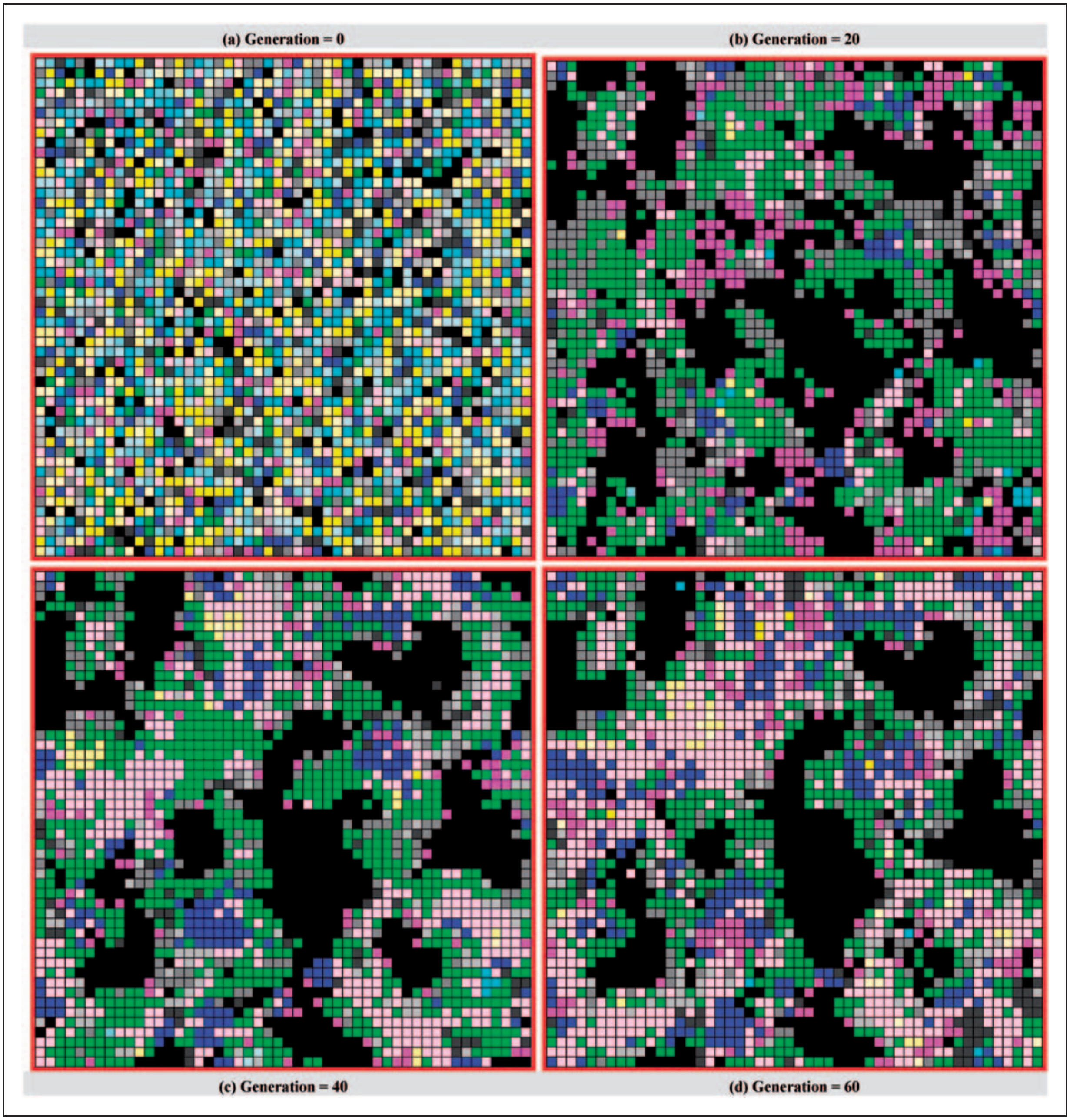

Figure 7. Spatial distribution in cellular automata of memory-1 deterministic strategic agents without self-awareness.

private interests. As stated earlier, when a majority of agents adopt that strategy, the entire community or society will likely fall into a cycle in which overall and individual private benefits rapidly decrease. In cellular automata, if the majority of an agent's adjacent nodes tend to adopt the same scoundrel strategy, then the agent in the center will be forced to adopt the same strategy for the sake of survival.

During the third stage (generations 11 to 20), agents wanting to counter the all-defection strategy tend to move toward a tit-for-tat strategy. In addition to confronting scoundrel strategy agents, this movement also supports cooperation with 'yes-men' (all-cooperation strategy agents) and other agents that also adopt the tit-for-tat strategy. Figure 7(b) illustrates a scenario in which agents who adopt the tit-for-tat strategy gradually increase in number and cluster in a manner that surrounds and restricts agents who adopt the all-defection strategy.

During the fourth stage (generations 21 to 40), the number of tit-for-tat strategy agents declines. Due to an asymmetry problem involving memories of previous encounters, tit-for-tat agents start to defect and stop trusting one another, resulting in less clustering over large areas. However, as shown in Figure 7(c), a certain number of tit-for-tat agents continue to surround all-defection agents to ensure that the latter group does not expand to the point of overwhelming tit-for-tat agents. Note also that as clusters of tit-for-tat agents start to break up and decrease in size, the number of agents adopting the win-stay, lose-shift strategy increases. Since win-stay, lose-shift agents do not have asymmetric memory problems regarding previous encounters (which increases the potential for promise-breaking), and since win-stay, lose-shift agents generally move toward mutual cooperation, their numbers and tendency to cooperate gradually increase.

During the fifth stage (generations 41 to 100), strategy evolution enters a state of 'dynamic stability.' Within clusters of win-stay, lose-shift agents, the number of agents who adopt an all-cooperation strategy gradually increases (Figure 7(d)). However, in reaction to this increase, some agents take advantage of the situation by reverting to the all-defection strategy, which reduces (or in some cases, eliminates) clusters of all-cooperation agents. The term 'dynamic stability' refers to the idea that this process enters a long period of repetition.

5. Results

Graphical representations of our simulations are shown in Figure 8. The Figure 8(a) social network consists of cellular automata; in Figure 8(b) we present results for a two-dimensional WS small-world network. All parameters are identical. As indicated by the pink average payoff curves in Figures 8(a) and 8(b), the full community or society clearly benefits when all agents possess the capacity for self-awareness, and a state of dynamic stability is achieved within a small number of generations. However, there can never be a situation in which all agents possess that capacity, therefore our focus was on determining the effects of adding a small number of self-aware agents to an otherwise unaltered environment. According to the green (10%) and yellow (30%) average payoff curves in the two figures, adding a small number of self-aware agents exerted a significant influence regardless of social network type. Specifically, they suppressed growth in the number of all-defection agents, prevented the initiation of a cycle in which all agents express betrayal and retaliation, and helped resolve conflicts between societal benefits and individual private interests so that cooperation could be accepted as mainstream behavior.

The average payoff curves in Figures 8(a) and 8(b) are similar because two-dimensional WS small-world networks contain many random long-distance shortcuts that reduce network separation. The main reason for a lack of strategic clustering is that these shortcuts (a) result in very low degrees of separation (approximately $\log N$, with N representing the total number of agents), and (b) significantly increase the level of complexity in terms of agent interactions and indirect influences. Note that the influence of a single strategy can result in increased evolutionary diffusion capacity and the increased containment of other agents. Combined, these factors speed up the movement toward dynamic stability.

5.1. Emergence of social behavior: Effects of four common strategies

The 'yes men' who follow an all-cooperation strategy are the most likely to be taken advantage of by agents who use tactics associated with an all-defection strategy – breaking promises, betraying opponents, and the like. In contrast, tit-for-tat agents find it easy to cooperate with all-cooperation agents and to attack all-defection agents. However, due to memory asymmetry problems regarding previous encounters, two tit-for-tat agents may collide and incidentally express behaviors such as breaking promises for an extended time period. Finally, win-stay, lose-shift agents tend to change their behaviors as soon as they acknowledge benefits from doing so. Note that these strategies can be used to categorize all other strategies.

We analyzed evolutionary dynamics and equilibrium for these four strategies using simulation results for different mixes of self-aware agents – 0% (Figures 9(a) and 9(d)), 100% (Figures 9(b) and 9(e)), and 10%

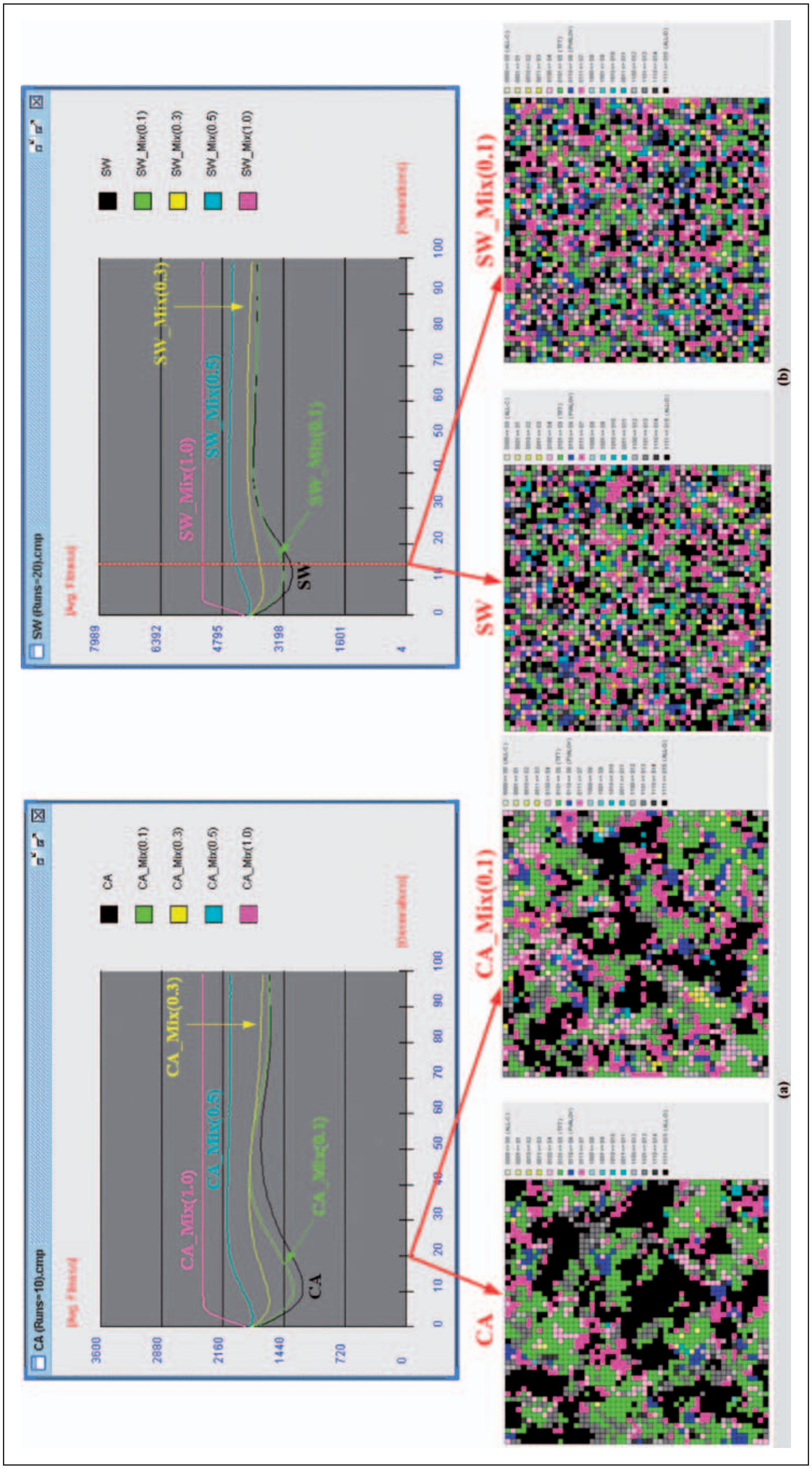

Figure 8. Comparison of results triggered by adding self-aware agents to (a) cellular automata and (b) a two-dimensional WS small-world network.

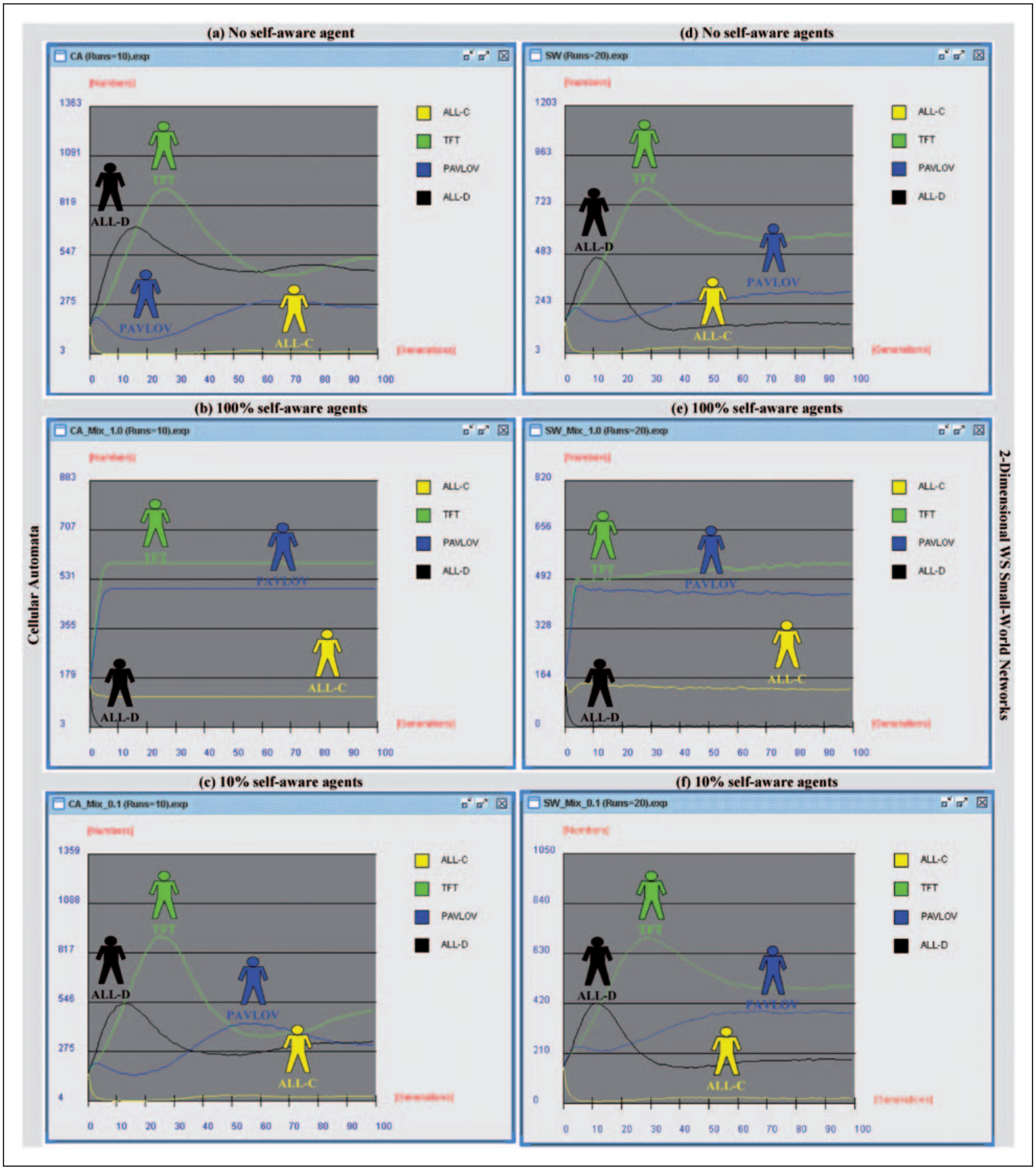

Figure 9. Evolutionary dynamics of four well-known prisoner's dilemma strategies.

(Figures 9(c) and 9(f)). Other than differences in social network type and self-aware agent percentages, all other parameter values were identical. According to Figures 9(a), 0% self-aware agents in a cellular automata resulted in roughly equal numbers of agents adopting each of the four strategies at the beginning of the simulation. After three generations, the number of agents adopting the all-defection strategy rapidly increased, and the number of agents adopting the all-cooperation or win-stay, lose-shift strategy slightly decreased. Tit-for-tat agents emerged when the number of all-defection agents reached a certain threshold. As described earlier, they confronted and suppressed all-defection agents and collaborated with all-cooperation and win-stay, lose-shift agents. After 20 generations, the number of tit-for-tat agents surpassed the number of all-defection agents, resulting in a sharp decrease in all-defection agents. The number of win-stay, lose-shift agents steadily increased after 30 generations, and after 60 generations the number of tit-for-tat agents fell below the number of all-defection agents. The increase in the number of all-defection agents had the result of reducing the number of win-stay, lose-shift agents. After 80 generations, the number of tit-for-tat agents once again surpassed the number of all-defection agents, and the simulated agent society entered a state of dynamic stability: the numbers of win-stay, lose-shift and all-cooperation agents did not change, and growth and decline in the numbers of all-defection and tit-for-tat agents balanced each other.

As shown in Figure 9(d), early evolutionary growth/decline rates for the four strategies in a two-dimensional WS small-world network with 0% self-aware agents were similar to those shown in Figure 9(a). After 30 generations, the number of all-defection agents reached a saturation point and remained at a fixed number that was clearly higher than their cellular automata counterparts. Due to the small-world network's characteristic of low degree of separation, the numbers of agents adopting each of the four strategies reached a state of dynamic stability between the fiftieth and sixtieth generations.

Figure 9(c) presents data on simulations involving cellular automata and a 10% addition of self-aware agents. Compared to Figure 9(a) (0% self-aware agents), the initial number of all-defection agents was not as great – a 150-agent difference. Figures 9(d) and 9(f) illustrate data for 0% and 10% additions of self-aware agents, respectively; here the difference in all-defection agent number was 60. As these figures show, there was a more significant delay in the emergence of betrayal/retaliation cycles when the percentage of self-aware agents was 10% rather than 100%. Note also that following the 10% addition of self-aware agents, the number of agents adopting a win-stay, lose-shift strategy surpassed the number of agents adopting the all-defection or tit-for-tat strategies, but after 20 generations the win-stay, lose-shift agents could not successfully resist the all-defection agents, even though their numbers increased. As a result, the number of win-stay, lose-shift agents started to decline to a stable level.

Figure 9(b) shows a cellular automata consisting of 100% self-aware agents. Since all-defection agents quickly discovered that their strategy was inappropriate for fulfilling social expectations, during early evolutionary stages all of those agents used their self-adjustment mechanisms to adopt other strategies to fulfill the expectations of adjacent agents. Starting at the third or fourth generation, the number of all-defection agents dropped to zero, and no new all-defection agents emerged for the rest of the simulation. The numbers of agents adopting the other three strategies also quickly stabilized without additional changes. Again, all parameters in Figures 9(b) and 9(e) were identical; the evolutionary dynamics of the four strategies in the two types of social networks were also virtually identical. The only significant difference was the presence of random long-distance shortcuts in the two-dimensional WS small-world network. Due to increased sensitivity, even small changes in a single agent's strategy were capable of influencing the entire society. However, due to the low degree of separation characteristic of WS small-world networks, a new state of dynamic stability was quickly reestablished.

6. Conclusion

Our proposed self-aware agent model features superego and ego personality traits, and includes an external learning mechanism and internal cognitive capacity. The model incorporates features taken from four research areas: AI, cognitive psychology, economics, and social/behavioral sciences. According to our results, the proposed model not only improves agent learning performance, but also provides a novel agent learning architecture. In terms of cognitive psychology, our self-aware agents can utilize personality traits to enhance their self-understanding and self-identity, thus promoting self-realization. The model also offers a novel approach to the iterated prisoner's dilemma: as long as a small number of self-aware agents are added to an IPD scenario, public good/private interest conflicts can be resolved, agent cooperation can be increased, and overall societal benefits can be enhanced. Finally, in terms of social/behavioral sciences, observing clustering behaviors allows for greater understanding of how self-awareness can influence evolutionary dynamics and equilibriums in artificial agent societies.

7. Funding

This research was supported in part by the Republic of China National Science Council (grant number NSC 99-2314-B-182-031), Chang Gung Memorial Hospital (grant number CMRPD290011), and Chang Gung University (grant number UEPRD280351). The funders had no role in study design, data collection and analysis, or preparation of the manuscript.

8. Conflict of interest statement

The authors declare no conflict of interest related to this work.

References

1. Baumeister RF. *The self in social psychology*. Philadelphia, PA: Psychology Press, 1999.
2. Aronson E, Wilson TD and Akert RM. *Social psychology*, 7th ed. Upper Saddle River, NJ: Prentice Hall, 2010.
3. Phillips AG and Silvia PJ. Self-awareness and the emotional consequences of self-discrepancies. *Person Soc Psychol Bull* 2005; 31(5): 703–713.
4. Peacocke C. Mental action and self-awareness. In: Cohen JD and McLaughlin BP (eds) *Contemporary debates in the philosophy of mind*. Oxford, UK: Blackwell Publishing, 2006.
5. Carver CS and Scheier MF. *Attention and self-regulation: a control-theory approach to human behavior*. New York: Springer-Verlag, 1981.
6. Noé E, Ferri J, Caballero MC, Villodre R, Sanchez A and Chirivella J. Self-awareness after acquired brain injury. *J Neurol* 2005; 252(2): 168–175.
7. Russell S and Norvig P. *Artificial intelligence: a modern approach*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2009.
8. Wooldridge M. *An introduction to multiagent systems*, 2nd ed. New York: John Wiley & Sons, 2009.
9. Ferber J. *Multi-agent systems: an introduction to distributed artificial intelligence*. New York: Addison-Wesley, 1999.
10. Goldspink C. Modelling social systems as complex: towards a social simulation meta-model. *J Artif Soc Social Simul* 2000; 3(2), http://jasss.soc.surrey.ac.uk/3/2/1.html.
11. Kawamura K, Noelle DC, Hambuchen KA, Rogers TE and Turkay E. A multi-agent approach to self-reflection for cognitive robotics. In: *Proceedings of the 11th International Conference on Advanced Robotics*. Coimbra, Portugal, 2003, pp.568–575.
12. Axelrod R. *The evolution of cooperation*. New York: Basic Books, 1984.
13. Axelrod R. *The complexity of cooperation: agent-based models of competition and collaboration*. Princeton, NJ: Princeton University Press, 1997.
14. Gilbert N. *Agent-based models*. Thousand Oaks, CA: Sage Publications, Inc, 2008.
15. Gilbert N and Troitzsch KG. *Simulation for the social scientist*, 2nd ed. New York: Open University Press, 2005.
16. Boero R and Squazzoni F. Does empirical embeddedness matter? Methodological issues on agent-based models for analytical social science. *J Artif Soc Social Simul* 2005; 8(4), http://jasss.soc.surrey.ac.uk/8/4/6.html.
17. Huang CY, Sun CT, Hsieh JL and Lin H. Simulating SARS: small-world epidemiological modeling and public health policy assessments. *J Artif Soc Social Simul* 2004; 7(4), http://jasss.soc.surrey.ac.uk/7/4/2.html.
18. Huang CY, Sun CT and Lin HC. Influence of local information on social simulations in small-world network models. *J Artif Soc Social Simul* 2005; 8(4), http://jasss.soc.surrey.ac.uk/8/4/8.html.
19. Sutton RS and Barto AG. *Reinforcement learning: an introduction*. Cambridge, MA: MIT Press, 2002.
20. Mitchell T. *Machine learning*. Boston, MA: McGraw-Hill, 1997.
21. Mitchell M. *An introduction to genetic algorithms*. Cambridge, MA: MIT Press, 1998.
22. Huang CY and Sun CT. parameter adaptation within co-adaptive learning classifier systems. *Lect Notes Comput Sci* 2004; 3103: 774–784.
23. Dobbyn C and Stuart S. The self as an embedded agent. *Minds Machines* 2004; 13(2): 187–201.
24. Markus H. Self-schemata and processing information about the self. *J Person Social Psychol* 1977; 35: 63–78.
25. Morris D. *The naked ape: a zoologist's study of the human animal*. New York: Delta Press, 1999.
26. Bem DJ. Self-perception theory. In: Berkowitz L (ed.) *Advances in experimental social psychology*. New York: Academic Press, 1972, Vol. 6, pp.1–62.
27. Smith ER and Mackie DM. *Social psychology*. New York: Worth Publishers, 2000.
28. Mead GH. *Mind, self, and society*. Chicago: University of Chicago Press, 1934.
29. Andersen SM and Chen S. The relational self: an interpersonal social-cognitive theory. *Psychol Rev* 2002; 109: 619–645.
30. Tice DM and Wallace HM. The reflected self: creating yourself as (you think) others see you. In: Leary MR and Tangney JP (eds) *Handbook of self and identity*. New York: Guilford Press, 2003, pp.71–88.
31. Wilson SW. State of XCS classifier system research. In: Lanzi PL, Stolzmann W and Wilson SW (eds) *Proceedings of learning classifier systems, from foundations to applications*. Berlin: Germany, 2000, pp.63–82.
32. Rasmusen E. *Games and information: an introduction to game theory*, 4th ed. New York: Wiley-Blackwell, 2006.
33. Aumann RJ. Game theory. In: Durlauf SN and Blume LE (eds) *The new Palgrave dictionary of economics*, 2nd ed. London: Palgrave Macmillan, 2008.
34. Nowak M and Sigmund K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature* 1992; 364(6432): 56–58.
35. Wu HC and Sun CT. What should we do before running a social simulation? *Social Sci Comput Rev* 2005; 23(2): 221–234.
36. Strachey J. *The standard edition of the complete psychological works of Sigmund Freud, Volume XIV (1914–1916): on the history of the psycho-analytic movement, papers on metapsychology and other works, ii-viii*. London: The Hogarth Press and the Institute of Psychoanalysis, 2001.
37. Myers DG and Ludwig T. *Psychology in modules (spiral-bound), visual concept reviews, psych inquiry and Scientific American reader to accompany Myers*. New York: Worth Publishers, 2007.
38. Bayne T and Pacherie E. Narrators and comparators: the architecture of agentive self-awareness. *Synthese* 2007; 159(3): 475–491.
39. Wheeler S, Petty R and Bizer G. Self-schema matching and attitude change: situational and dispositional determinants of message elaboration. *J Consum Res* 2005; 31(4): 787–797.

40. Mui L, Mohtashemi M and Halberstadt A. Notions of reputation in multi-agent systems: a review. In: *Proceedings of First International Joint Conference on autonomous agents and multi-agent systems*. Bologna, Italy, 2002, pp.280–287.
41. Sabater J and Sierra C. Reputation and social network analysis in multi-agent systems. In: *Proceedings of First International Joint Conference on autonomous agents and multi-agent systems*. Bologna, Italy, 2002, pp.475–482.
42. Huynh T, Jennings N and Shadbolt N. An integrated trust and reputation model for open multi-agent systems. *Auton Agents Multi-Agent Syst* 2006; 13(2): 119–154.
43. Watts DJ and Strogatz SH. collective dynamics of 'small-world' networks. *Nature* 1998; 393(6684): 440–442.

Chung-Yuan Huang received his MS in computer information and science (2000) and his PhD in computer science (2005), both from the National Chiao Tung University, Taiwan. He is currently an Associate Professor in the Department of Computer Science and Information Engineering and a member of the Research Center for Emerging Viral Infections at Chang Gung University, Taiwan. His research interests include complex adaptive networks and systems, agent-based modeling and simulation for social science research, and computational epidemiology.

Sheng-Wen Wang received his BS and MS degrees in computer science from the Tamkang and the National Tsing Hua University, Taiwan, in 1999 and 2001, respectively. He is currently a PhD candidate in the Department of Computer Science, National Chiao Tung University. His current research interests include social simulations and multi-agent systems.

Chuen-Tsai Sun received his BS degree in electrical engineering (1979) and his MA in history (1984) from National Taiwan University. He earned his PhD in computer Science (1992) from the University of California at Berkeley. He joined the faculty at National Chiao Tung University in 1992. He is currently a joint Professor of Department of Computer Science and Graduate Institute of Education, National Chiao Tung University, Taiwan. His research interests include digital games, digital learning, social networking, and simulation modeling in the social sciences.