Emergence and Stability of Social Conventions in Conflict Situations

Toshiharu Sugawara
Department of Computer Science and Engieering
Waseda University, Japan
sugawara@waseda.ac.jp

Abstract

We investigate the emergence and stability of social conventions for efficiently resolving conflicts through reinforcement learning. Facilitation of coordination and conflict resolution is an important issue in multi-agent systems. However, exhibiting coordinated and negotiation activities is computationally expensive. In this paper, we first describe a conflict situation using a Markov game which is iterated if the agents fail to resolve their conflicts, where the repeated failures result in an inefficient society. Using this game, we show that social conventions for resolving conflicts emerge, but their stability and social efficiency depend on the payoff matrices that characterize the agents. We also examine how unbalanced populations and small heterogeneous agents affect efficiency and stability of the resulting conventions. Our results show that (a) a type of indecisive agent that is generous for adverse results leads to unstable societies, and (b) selfish agents that have an explicit order of benefits make societies stable and efficient.

1 Introduction

Norms (or social laws) and social conventions in agent societies have received much attention in terms of how they facilitate coordination and conflict resolution. A norm can be considered a restriction on a set of actions available to agents, and a social convention is a special type of norm that restricts the agents' behavior to a particular strategy [Shoham and Tennenholtz, 1997]. According to [Pujol *et al.*, 2005], a social convention is a regularity of agents' behaviors as a result of being a solution to a recurrent coordination problem. Because all agents in a society are expected to follow these norms and conventions, they can achieve coordination and conflict resolution with no explicit communications. Thus, they can significantly reduce the cost of resolving coordination problems, especially in environments where a huge number of agents work together and the situations requiring coordination and conflict resolution frequently occur, such as service computing in Internet and sensor network applications. Thus, the emergence of conventions and their stability in society are the main concerns in literature on a multi-agent systems.

A number of studies have addressed conventions and norms. One important study from this viewpoint is [Shoham and Tennenholtz, 1992], which first addressed the issue of formalizing the norms and conventions and synthesizing them during the design process. However, it is not easy to develop all of the useful conventions in the design stage. Thus, [Shoham and Tennenholtz, 1997] have addressed the emergence of conventions in a multi-agent context and investigated how they emerge under a number of rationality assumptions using coordination games that have a single potential convention (equilibrium). Subsequently, many researchers have studied the emergence of conventions; For example, [Walker and Wooldridge, 1995] proposed other rationalities that can result in the emergence of conventions. Recently, the emergence of norms and conventions have been investigated from game theoretic approaches where all agents select actions based on their own payoff matrix and identify the best actions from their own viewpoint using reinforcement learning (e.g. [Savarimuthu *et al.*, 2008; Sen and Airiau, 2007]). Furthermore, for coordination games having multiple equilibria, [Mukherjee *et al.*, 2008] showed that all agents develop a norm/convention in order to select one of the equilibria via reinforcement learning and local interactions, and [Pujol *et al.*, 2005] have investigated the role of the agent's network structure in the emergence of social conventions. However, these studies mainly focused on the emergence of conventions of coordination games.

However, agents in actual applications often encounter conflict situations expressed as a non-coordination game. This means that it does not have an obvious equilibrium. Furthermore, if they fail to resolve the conflict, the conflict situation still remains, and thus, the game is iterated. Hence, the emergent social conventions should resolve the conflicts as quickly as possible from the social perspective; that is, they can reduce the number of game iterations as well as enlarge the payoffs of individual agents.

Stability of emergent conventions is another important issue: If agents continuously use an emergent convention, but it is altered to another one, the results from the convention may become undesirable. However, if agents are conservative and never change the convention, the agents will not be able to adapt to the open environments.

The aim of this paper is to investigate the question of whether conventions that lead to efficient conflict resolution can emerge in competitive and conflict situations using reinforcement learning, even though agents act and learn according to their own payoff matrices. We expressed this situation as a Markov game whose best policy is non-trivial for all agents since the payoff matrices in the adversary agents are unknown. We also focus on the stability of the conventions if agents continuously use them. For this purpose, we introduce a number of payoff matrices that characterize the agents' decision, in order to understand how the characteristics affect the emergence of conventions, the resulting societies, and their stability. We analyze the resulting societies in detail, in order to understand why the resulting efficiencies differ and why the emergent conventions are (un)stable.

Another feature of our research is that an agent in the society plays this game with an anonymous agent, and it learns the policy from the results of games with different agents, as in the *social learning* of [Mukherjee *et al.*, 2008]. Thus, no agent has prior knowledge about their adversary, and each has to act on the basis of only the strategy it has learned so far. Hence, agents have no choice but to use conventions.

This paper is organized as follows: first, we discuss the model of the game by describing the conflicts and the issues to be addressed. Next, we introduce the agents that are characterized by their payoff matrices. Then we describe the experiments that show what agent characteristics result in conventions for social efficiency and their stability. Our results indicate that agents having explicit orders of actions can achieve the emergence of social conventions that are stable and can lead to an efficient society, whereas those that are not willing to give the other an advantage (that is, negative local payoffs) cannot develop such conventions. We also show that a type of indecisive agent that is generous with adverse results can achieve the emergence of conventions for an efficient society but that is unstable.

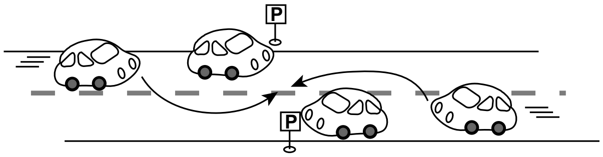

Figure 1: Narrow road game.

2 Model and Problem

2.1 Narrow Road Game in Agent Society

To clearly describe a conflict situation, we consider a modified version of the *narrow road game* (MNR game) [Moriyama and Numao, 2003] in which car agents encounter the situation shown in Fig. 1. This is a two-player game, more precisely a sort of Markov game or stochastic game [Moriyama and Numao, 2003; Littman, 1994], expressed by the following payoff matrix where the agents take one of two actions, p (proceed) or s (stay):

$$\begin{array}{cc} & \begin{array}{cc} p & s \leftarrow \text{Actions of the adversary agent.} \end{array} \\ \begin{array}{c} p \\ s \end{array} & \left(\begin{array}{cc} -5 & 3 \\ -0.5 & 0 \end{array}\right) \end{array} \qquad \text{(M1)}$$

Suppose that the joint action is denoted by (m_l, m_a), where m_l and m_a are the actions of local and adversary agents, respectively. If the local and adversary agents take the same actions, that is, their joint action is (p, p) or (s, s), the game does not end and they will play a second round; the game is iterated until they take different actions.

The agents having matrix (M1) receives -5 (maximum penalty) if their action is (p, p) because they have to go back to escape the deadlock. However, (s, s) does not induce any benefit or penalty because no progress occurs (later, we introduce a small penalty for (s, s)). The action pair (s, p) induces a small penalty (-0.5) because the adversary agent has priority, but the local agent can proceed right after that. Of course (p, s) has the maximum benefit, since the local agent has priority over the adversary agent.

The characteristic of this game is the unbalanced penalties for (p, p) and (s, s), and we should emphasize that this kind of situation often occurs in conflicts; if one of the conflicting agents moves without considering the other agents, it may result in a significant penalty. However, if all agents concerned take the wait-and-see strategy, nothing happens; this leads to zero or small penalties and another round of the game.

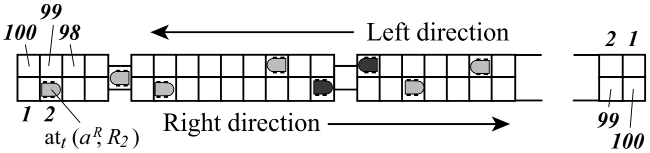

Figure 2: Modified narrow road game.

We consider that agents in two parties A_L and A_R, which are the disjoint sets of agents, play the MNR game. We also assume that $A = A_L \cup A_R$ is the society of agents. The two-lane road, as shown in Fig. 2, is one in which agents in A_L (A_R) move forward in the left (right) lane. Each position on the road is represented by a cell denoted by L_i or R_j, where i and j ($1 \leq i, j \leq l$) are the indices in Fig. 2 and l is the length of the road (so $l = 100$ in Fig. 2). The notation $\text{at}_t(a_i^D, D_j)$ means that agent $a_i^D \in A_D$ is at D_j at time t, where $D = L$ or R. No two agents can be at the same cell. Agent $a_i^D (\in A_D)$ at D_j moves forward to D_{j+1} (if $j \neq l$) or D_1 (if $j = l$) every time if no agent is in the forward position. The road has a number of narrow parts where left and right lanes converge into one. The narrow part is expressed by a single cell so only one agent can be in it. Hence, if L_k is a narrow part, then $L_k = R_{l-k+1}$.

In this environment, two agents $a_i^L \in A_L$ and $a_j^R \in A_R$ play the narrowroad game when they are on ether side of a narrow part and must avoid collision. More precisely, if, at time t,

(1) $\exists k, \text{at}_t(a_i^L, L_{k-1}) \wedge \text{at}_t(a_j^R, R_{l-k})$,
(2) L_k and R_{l-k+1} are the same narrow part, and
(3) no agent is in $L_k (= R_{l-k+1})$,

then a_i^L and a_j^R begin to play the MNR game. For example, the darkly colored agents in Fig. 2 will play the game. A game started at t is denoted by G_t. Agent a can proceed at

$t+1$, that is, $\mathrm{at}_{t+1}(a, L_k)$, if its action is p, and the adversary agent takes action s and G_t ends (and the adversary agent will be able to proceed at $t+3$ if the above conditions do not hold).

Resource conflicts similar to the MNR game can be found in real applications. An obvious example is robots moving about in a room. They may have to pass through narrow spaces, such as through a door or a space between a desk and a bookshelf. Also in Internet services, such as grid computing, cloud computing, and service computing, where a great deal of tasks (or service components) are requested by many different agents, a number of tasks are sometimes requested simultaneously, and this can slow down the server, cause tasks to be dropped or, in rare cases, cause thrashing/livelock due to a lack of resources. In such a case, the agents have to cancel/stop the current jobs and request/resume them again. In our model, agents in a social party correspond to those requesting a certain task or service component as a part of the larger task and the payoff matrix expresses the original strategy of the agents when the conflicts occur. Conflicts may occur with any type of agent, some of which may have just been added to the systems, so the appropriate social behaviors cannot be decided *a priori*. Thus, they have to learn appropriate social conventions so that they can resolve conflicts as soon as possible and with less effort even though they will receive an initial small penalty.

2.2 Emergence of Conventions

We investigated how agents learn the conventions for the MNR games by reinforcement learning and how their society becomes efficient as a result of the emergent conventions. We shall use the term *policy*, which is used in reinforcement learning literature, to express what action will be taken in each state. Thus, we can say that the conventions are learned policies each of which is common in each social party A_L or A_R and that are *consistent* with each other. Thus, two conventions in A_R and A_L are called *joint conventions*. Note that consistent conventions mean that the joint actions induced by the joint conventions can immediately resolve a conflict.

For each game, agent $a \in A$ receives a payoff v as a positive or negative reward. If a cannot resolve the conflict (that is, the game does not end), a has to play the MNR game again with the same adversary. The MNR game can be expressed as the Markov process shown in Fig. 3. Note that, in this figure, $S(= W_0)$ and T are the start and terminal states, and W_n indicates that the agent plays the $n+1$-th MNR game and so has already come to a standstill n times.

Unlike the dotted-line nodes in Fig. 3, after the start state, the agent does not enter different states depending on the actions of the adversary agents (this is the same as the original narrow-road game [Moriyama and Numao, 2003]). This means that the agent does not change its policy in accordance with the previous action of the adversary agent. Rather, the policy depends only on its local states. As mentioned before, the agent plays the MNR game with an anonymous agent that may have inconsistent policies (at least, this is true before conventions emerge). Nevertheless, agents have to resolve a conflict with less effort using social conventions. Agents cannot use prior knowledge about the adversary agents for the emergence of conventions.

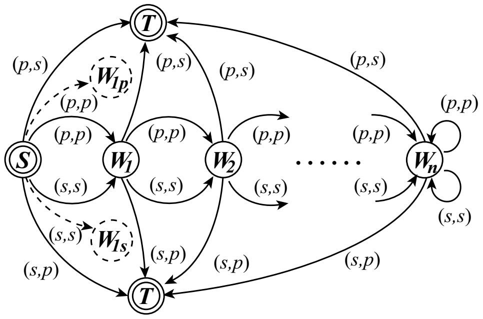

Figure 3: State transitions of modified narrow road game.

We also note that the games may be iterated any number of times, but that agents go back to state W_n if the game is iterated more than n times in this model, as shown in Fig. 3. In the following experiments we set $n = 5$ because agents rarely entered W_5 after sufficient learning steps.

All the agents learn Q values for all states by using,

$$\begin{aligned} Q(W_n, m) \quad &\leftarrow \quad (1-\alpha)Q(W_n, m) \\ &+ \quad \alpha[r(W_n, m) + \gamma \max_{m'} Q(W_{n+1}, m')] \end{aligned} \quad (1)$$

where m is a possible action, α is a learning rate, γ is a discount factor, and the reward is a payoff defined in the payoff matrix. Agents take actions based on the Q value for each state by using the ε-greedy strategy.

We expect that agents in A_L (or A_R) will learn common conventions and that the conventions of A_L and A_R may be different from each other but consistent. However, by introducing a state variable indicating the direction of an agent into the Markov model, it is likely that all agents in A can devise the same conventions that take different actions for different directions. However, we do not merge the parties since we intend to investigate how the efficiency of the emergent societies is affected by characteristics of agents.

2.3 A Variety of Payoff Matrices

We introduce a number of payoff matrices that characterize the agents in order to investigate their effect on the emergent conventions:

$(M2)$ Moderate
$$\begin{pmatrix} -5 & 3 \\ 0.5 & 0 \end{pmatrix}$$

$(M3)$ Selfish
$$\begin{pmatrix} -5 & 3 \\ -0.5 & -0.5 \end{pmatrix}$$

$(M4)$ Generous
$$\begin{pmatrix} -5 & 3 \\ 3 & -0.5 \end{pmatrix}$$

$(M5)$ Self-centered
$$\begin{pmatrix} -5 & 3 \\ -5 & -0.5 \end{pmatrix}$$

Note that we call an agent characterized by matrix M1 *normal*. An agent has only one payoff matrix.

Matrix M2 characterizes a *moderate* agent whose payoff of (s, p) is 0.5 (positive); it may be able to proceed the next time. The *selfish* (or *self-interested*) agent is characterized by M3, which has a positive payoff only when it can proceed the next time. (Joint action (s, s) also induces a small penalty because it is a waste of time). The *generous* agent defined by M4 does not mind if its adversary proceeds first (it can proceed the next time if the game is over). This matrix defines the coordination game and has two obvious equilibria [Mukherjee *et al.*, 2008] if this is a singe shot game. The *self-centered* agent characterized by M5 is only satisfied when it can proceed and is very unhappy if the adversary goes first. Matrix M5 has the obvious best action p if the game is not iterative.

3 Experiment — Emergent Conventions from Payoff Matrices

3.1 Experimental Setting

We assume that the populations of both parties $|A_L|$ and $|A_R|$ are 20, the road length l is 100, and there are four narrow parts along the road (the positions are random). All agents in A_L (A_R) are randomly placed on the left (right) lane except for the narrow parts. We also define a small probability β whereby agent a does not proceed with probability β even if its next forward position is empty. Parameter β avoids situations where no games occur.[1] We set $\alpha = 0.05$, $\gamma = 0.95$, $\varepsilon = 0.05$, and $\beta = 0.001$. The data shown in this paper are the average values of 1000 trials.

3.2 Improvement of Social Efficiency

We assume that A_R and A_L consist of homogeneous agents. The first experiment investigated how reinforcement learning over time shortens the time required for agents to go round (that is, the time required to come back to the start position; this is called the *go-round time*). All the experimental results shown in this paper are the average values of the data calculated through 1000 trials in the simulation environment. We compared the average go-round times (AGRT) of the societies. The results are shown in Fig. 4, where the AGRT values are plotted every 1000 times from $t = 0$ to 200000.

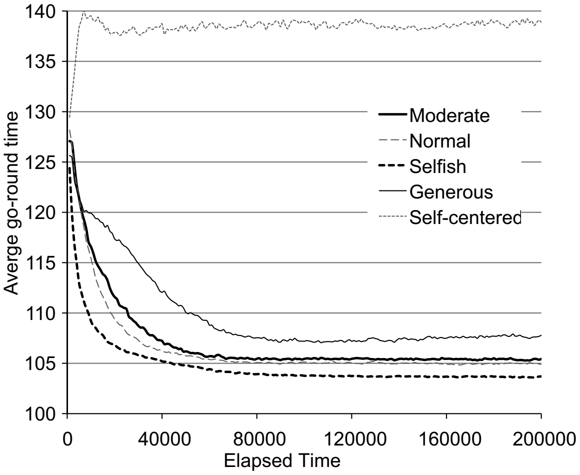

Figure 4: Average go-round time (AGRT).

This figure indicates that the AGRT values become smaller in all societies except the self-centered one. Because a smaller AGRT means that conflicts can be resolved more quickly, we can say that the society becomes more efficient through reinforcement learning.

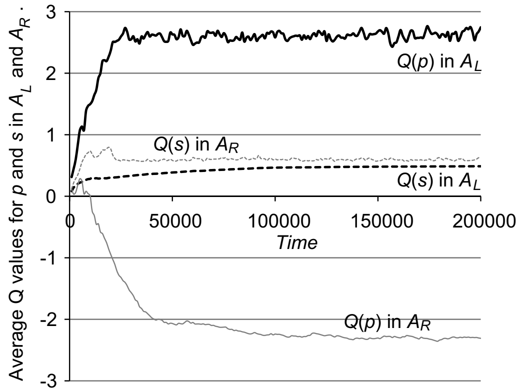

Figure 5: Average Q values in state S for actions p and s in A_L and A_R with moderate agents.

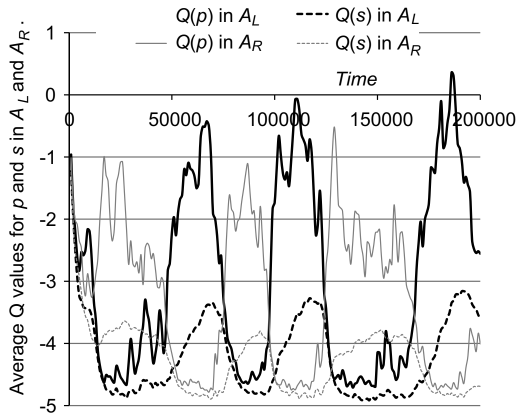

Figure 6: Average Q values in state S for actions p and s in A_L and A_R with self-centered agents.

To check whether or not social conventions really emerge, we looked into the changes in the average Q-values in state S of A_L and A_R of a typical example from the 1000 trials in the moderate society. The results are shown in Fig. 5. The agents in A_L had a higher average Q-value for p than for s, whereas those in A_R had opposite values. In fact, for $t \geq 22000$, all agents in A_L acquired a policy selecting p, and all those in A_R acquired one selecting s. These conventions are consistent with each other in that they indicate less effort in resolving conflicts and thus result in high efficiency. Although in this example, the conventions that prioritize A_L over A_R, (these are called *left-priority conventions*) emerge, in general, left- and right-priority conventions emerge with equal probability in the societies of normal, moderate, selfish, and generous agents.

[1] For example, games never occur if all the agents are evenly spaced apart.

No consistent conventions emerged in the self-centered society, however. Figure 6 plots the changes in the average Q-values in S over time using the same trial (identical random seed). The Q values periodically vary, and almost all agents in A prefer p; the society becomes very competitive, and the AGRT values never become smaller.

3.3 Characteristics of Emergent Conventions

Left- or right-priority conventions usually emerge except in the self-centered society. However, Fig. 4 indicates that the convergence speed and resulting efficiency depends on the characteristics of the agent. To account for these differences, we analyzed the emergent conventions in detail.

First, let us denote the prior party by A_P, and the other party by A_N, where $P, N = L$ or R and $P \neq N$ for each trial. We define $\mathcal{N}^t(m, W, \mathcal{P}) (\leq 20)$ to be the number of agents in party $\mathcal{P}$ that have the policy to select action m in state W at time t, where $m = p$ or s, $W = S, T$ or W_n, and $\mathcal{P} = A_L, A_R, A_P$, or A_N. If $t = 300000$, the superscript t is omitted (we stopped the simulation at $t = 300000$ because of limited resources, but we believe that this period was long enough to get a clear picture of the emergent conventions and their changes over time). Note that $\mathcal{N}^t(p, W, \mathcal{P}) + \mathcal{N}^t(s, W, \mathcal{P}) = 20$.

In the ideal case, the learned policies evolve into joint conventions in which all agents in A_P select p and others select s; that is,

$$\mathcal{N}^t(p, S, A_P) = 20 \wedge \mathcal{N}^t(s, S, A_N) = 20. \qquad (2)$$

The joint conventions of this type are called *efficient.* Conversely, if agents in either party do not evolve conventions, that is, $\mathcal{N}^t(p, S, A_L) \neq 0 \wedge \mathcal{N}^t(p, S, A_R) \neq 0$, these rules (not conventions) are called *scrambled* because a number of agents have policies resulting in an inconsistent joint action (p, p).

We found that there are intermediate cases between the efficient and scrambled cases, for example, $\mathcal{N}^t(s, S, A_N) = \mathcal{N}^t(s, W_1, A_N) = 20$. This occurs because the second round of the game is independent from the first round. In such a case, if $a_P \in A_P$ takes s (this occurs with probability ε) and enters the second round of the MNR game, a_P is still privileged in W_1. This increases $Q(S, s)$ of a_P. Thus, conditions

$$\begin{array}{rcl} 20 - \mathcal{N}^t(s, S, A_P) & = & \mathcal{N}^t(p, S, A_P) < 20 \text{ and} \\ \mathcal{N}^t(p, W_1, A_P) & = & 20 \end{array} \qquad (3)$$

hold for sufficiently large t. This increases the chance of a joint action (s, s) in the first round but, of course, it is a waste of time. In such a case, agents in A_N likely stay (do not proceed) until $t + 2$ of game G_t if the adversary agent does not select p. Such joint conventions are called *2-iterative* in this paper.

Similarly, the following cases sometimes occur:

$$\begin{array}{rcl} \mathcal{N}^t(s, W_0, A_N) = & \ldots & = \mathcal{N}^t(s, W_k, A_N) = 20 \\ \mathcal{N}^t(s, W_{k+1}, A_N) & \neq & 20 \\ \mathcal{N}^t(p, W_k, A_P) & = & 20 \end{array} \qquad (4)$$

where $S = W_0$. Here, agents in A_N likely select the joint action (s, s) k times; thus, they likely stay $k + 1$ times. When the society's actions lead to the emergence of conventions by which agents in A_N likely wait $k + 1$ times if the game is iterated k times, the joint conventions are called *potentially k-iterative* or simply *k-iterative*.

Table 1: Numbers of emergent convention types.

Types of agents	$\mathcal{C}_{\mathit{eff}}$	$\mathcal{C}_{\mathit{float}}$	$\mathcal{C}_{\mathit{scram}}$
Moderate	638	362	0
Selfish	851	121	28
Generous	64	924	12
Self-centered	0	0	1000

Let us define notations to classify the 1000 trials of the experiments. $\mathcal{C}^t_k$ is the number of trials that evolved k-iterative conventions at t. Since the efficient conventions are 1-iterative, $\mathcal{C}^t_1$ is also denoted by $\mathcal{C}^t_{\mathit{eff}}$. If $t = 300000$, the superscript t is omitted hereafter. Similarly, we define $\mathcal{C}^t_{\mathit{scram}}$ as the number of trials that led to the scrambled conventions. $\mathcal{C}^t_{\mathit{trans}}$ is the number of transient trials in which the emergent conventions are neither k-iterative $(k \geq 1)$ nor scrambled. The transient trials seem to be on the way to the k-iterative conventions. Finally, the conventions that are neither efficient nor scrambled are called *floating* because the policies that emerged as conventions in A_P for state S may vary, and thus, the chance of game 'iterations increases (see Condition (3)). The number of trials that resulted in the evolution of floating conventions is $\mathcal{C}^t_{\mathit{float}} = \sum_{k \geq 2} \mathcal{C}^t_k + \mathcal{C}^t_{\mathit{trans}}$. Finally, we redefine the priority conventions. If the conventions are not scrambled, at least $\mathcal{N}^t(s, S, A_N) = 20$. If $N = R$ $(N = L)$, the conventions are called left-priority (right-priority) at t.

Table 1 lists the values of $\mathcal{C}_{\mathit{eff}}$, $\mathcal{C}_{\mathit{float}}$ and $\mathcal{C}_{\mathit{scram}}$. These results explain the differences in the AGRT values in Fig. 4, because $\mathcal{C}_{\mathit{eff}}$ of the selfish society is higher than that of the moderate society, and $\mathcal{C}_{\mathit{eff}}$ of the generous society is much lower than those of the others.

3.4 Stability of Emergent Conventions

To explore the temporal transition of emergent conventions, we show the changes in $\mathcal{C}^t_{\mathit{eff}}$, $\mathcal{C}^t_{\mathit{float}}$, $\mathcal{C}^t_{\mathit{scram}}$, $\mathcal{C}^t_k$, and $\mathcal{C}^t_{\mathit{trans}}$ over time in moderate and generous societies, in Fig. 7. These graphs suggest that the stability of emergent conventions in the moderate society and the instability of those in the generous society. From Fig. 7(b), we can see that $\mathcal{C}^t_{\mathit{eff}}$ decreases and $\mathcal{C}^t_{\mathit{float}}$ increases. Thus, efficient conventions gradually change into floating conventions over time. Figures 7(c) and (d) depict more detailed data. In the moderate society, k-iterative conventions are invariant, but in the generous society, the efficient conventions gradually turn into transient ones and then into k-iterative conventions ($k \geq 2$). Note that $\mathcal{C}^t_{\mathit{eff}} + \mathcal{C}^t_{\mathit{trans}}$ start to decrease after $t = 120000$. Actually, if we review Fig. 4 closely, we can see that AGRT becomes slightly worse after $t = 120000$. In the generous society, agents have two obvious policies and no clear preference between them. This prevents the AGRT values from improving. We also note that the emergent conventions in the selfish society were also stable whereas a small number of scrambled conventions were

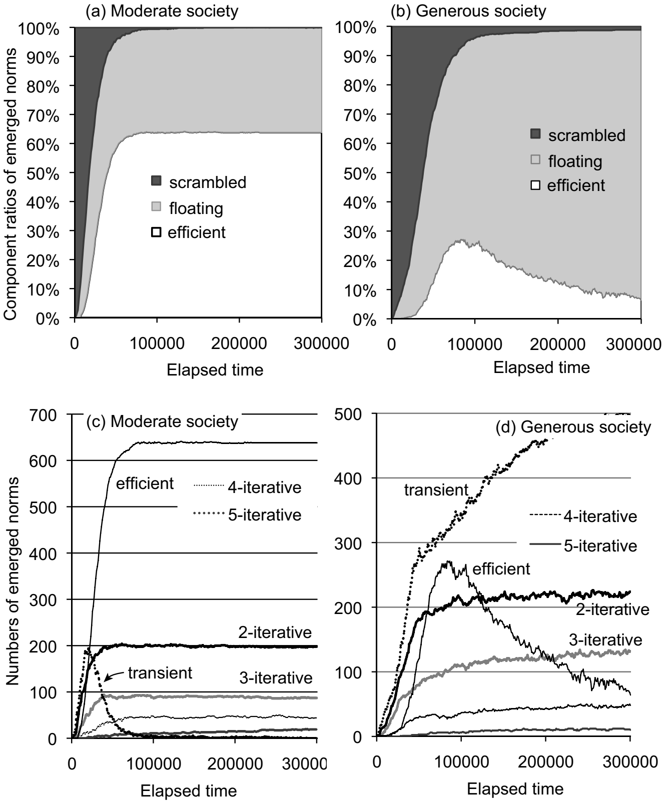

Figure 7: Ratios and numbers of emergent convention types over time.

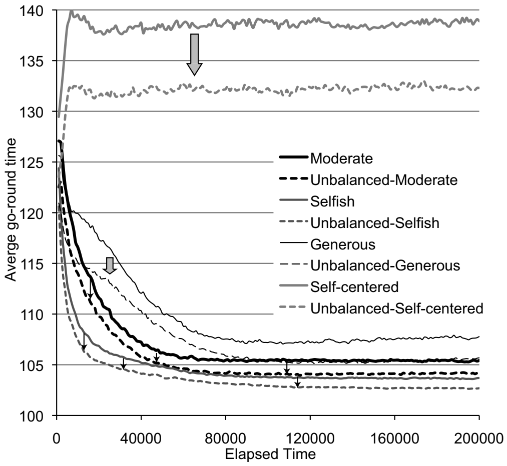

Figure 8: Average go-round times over times (unbalanced populations).

Table 2: Percent ratios of emergent societies (unbalanced populations).

Types of agents	Same populations		Unbalanced populations	
	$\mathcal{R}_L$	$\mathcal{R}_R$	$\mathcal{R}_L$	$\mathcal{R}_R$
Moderate	50.2%	49.8%	56.1%	43.9%
Selfish	50.8%	49.2%	63.0%	37.0%
Generous	48.9%	51.1%	73.8%	26.2%

emerged (Table 1.

Let us discuss why the conventions of indecisive agents, like the generous ones, are unstable. Agent a_N in A_N sometimes selects p (with probability ε), which results in the maximum penalty, so $Q(S,p)$ in a_N decreases. This behavior accords with the emergent convention in A_N. However, the adversary agent a_P also receives the maximum penalty and $Q(S,p)$ decreases. Eventually, a few agents in A_P begin to use the policy that selects s at S. On the other hand, the joint action (s,s) receives a relatively small penalty and a_P may be able to receive the benefit in the second round. The emergent efficient conventions in A_P then start to turn to floating conventions. Agents in the generous society have no clear opinion, and this makes the society unstable and inefficient. Thus, having a preference is important for maintaining conventions, although the best payoffs are not always realized. Note that this discussion explicitly indicates a difference between our results and those of [Mukherjee *et al.*, 2008].

3.5 Unbalanced Populations

The second experiment investigates how the different populations of parties affect the resulting conventions. We decreased the population of only A_R to 18. Note that the population difference means different chances in the games of the individual agents.

Figure 8 shows the AGRT values over time for the moderate, selfish, generous and self-centered society. The AGRT values for the same population ($|A_L| = |A_R| = 20$) are also indicated for comparison. These graphs show that the AGRT values become smaller. The reason is obvious; the chance of encounters is lower because of the smaller population.

More importantly, we are curious about which parties are prioritized. Let us denote $\mathcal{C}_L$ ($\mathcal{C}_R$) as the number of trials in which A_L (A_R) has priority. Table 2 compares their ratios, that is, $\mathcal{R}_L = \mathcal{C}_L/(\mathcal{C}_L + \mathcal{C}_R)$ and $\mathcal{R}_R = \mathcal{C}_R/(\mathcal{C}_L + \mathcal{C}_R)$, in all types of societies.

Table 2 indicates that party A_L has more priority than A_R in the unbalanced population cases, whereas they have the almost identical probabilities to be prioritized if they have the same populations. This result shows that bigger societies are likely to have priority and is preferable because more agents are prioritized.

3.6 Heterogeneous Agents

Finally, we explored the effect of adding a small number of heterogeneous agents to a party. Because the self-centered agents resulted in a quite different society (cf. Figs. 4 5 and 6), we replace two agents in A_R with self-centered agents in the moderate, selfish and generous societies, so that only A_R was heterogeneous. Note that $|A_L| = |A_R| = 20$.

There is no significant difference in the AGRT values from those in Fig. 4 (so the corresponding figure omitted), but the numbers of the prioritized parties $\mathcal{R}_L$ and $\mathcal{R}_R$ are quite dif-

Table 3: Percent ratios and numbers of prioritized societies and emergent conventions.

Types of agents	$\mathcal{R}_L$	$\mathcal{R}_R$	$\mathcal{C}_{\mathit{eff}}$	$\mathcal{C}_{\mathit{float}}$	$\mathcal{C}_{\mathit{scram}}$
Moderate	70.0%	30.0%	30.8%	35.8%	33.4%
Selfish	72.8%	27.2%	61.5%	14.5%	24.0%
Generous	62.1%	37.9%	24.3%	49.5%	26.2%

ferent. Table 3 lists the results together with the ratios of $\mathcal{C}_{\mathit{eff}}$, $\mathcal{C}_{\mathit{float}}$, and $\mathcal{C}_{\mathit{scram}}$ to the number. $\mathcal{C}_{\mathit{scram}}$ is larger than in Table 1 for any type of society. Particularly, scrambled conventions were not emerged in the homogeneous moderate society (Table 1) but scrambled conventions were emerged in the heterogeneous moderate society more than those in the heterogeneous selfish society.

By comparing $\mathcal{R}_L$ and $\mathcal{R}_R$, we can see that the homogeneous parties, A_L is likely to have priority over the heterogeneous ones, A_R. However, this result is counter-intuitive, because the self-centered agents seem to prefer p.

To explain this phenomenon, we must take into account three facts. First, all the agents almost randomly select their actions at the beginning of the experiments. Moreover, before the conventions emerge, agents in the adversary party may have different policies, so their actions also seem almost random. Second, if the adversary agents select actions randomly, the self-centered agents become unhappier than the other types of agents, because the average payoff values in the self-centered matrix is much lower than those of the others. Hence, the self-centered agents receive a larger penalty when they fail to resolve their conflicts or the adversary agent has priority. Finally, we also have to consider the results of the experiments on the unbalanced populations because the situations are quite similar; non self-centered agents in A_R play the MNR game with those in A_L and never play with self-centered agents. These considerations mean that the A_L has priority over A_R before the conventions emerge. In fact, Fig. 9(a), which shows $\mathcal{C}^t_L$ and $\mathcal{C}^t_R$, indicates that the left-priority conventions emerge with higher probability in the moderate society with two self-centered agents.

However, the situations are quite different after the conventions emerged, as shown in Fig. 9(a). This figure suggests that the emergent left-priority conventions are gradually destroyed by two self-centered agents after $t = 100000$. If A_L has priority over A_R, the self-centered agents in A_R usually select s, and this results in the maximum penalty. Thus, $Q(S, s)$ gradually falls, and they begin to select p. This confuses agents in the other society. Figures 9(c) and (e) are the detailed graphs of the emergent conventions. They clearly indicate that efficient conventions decrease, and they also show that $\mathcal{C}^t_{\mathit{scram}}$ gradually increase; this cannot be observed in homogeneous cases (Fig. 7).

On the other hand, the selfish society is relatively robust against confusion from self-centered agents. Figures 9 (b) (d) and (f) show that $\mathcal{C}^t_{\mathit{eff}}$ rarely decreased and $\mathcal{C}^t_{\mathit{scram}}$ never increased. Note that Fig. 9(f) has no 4- and 5-iterative conventions. Thus, characteristic like selfish agents is quite important to evolve efficient conventions for efficient society and to sustain the conventions.

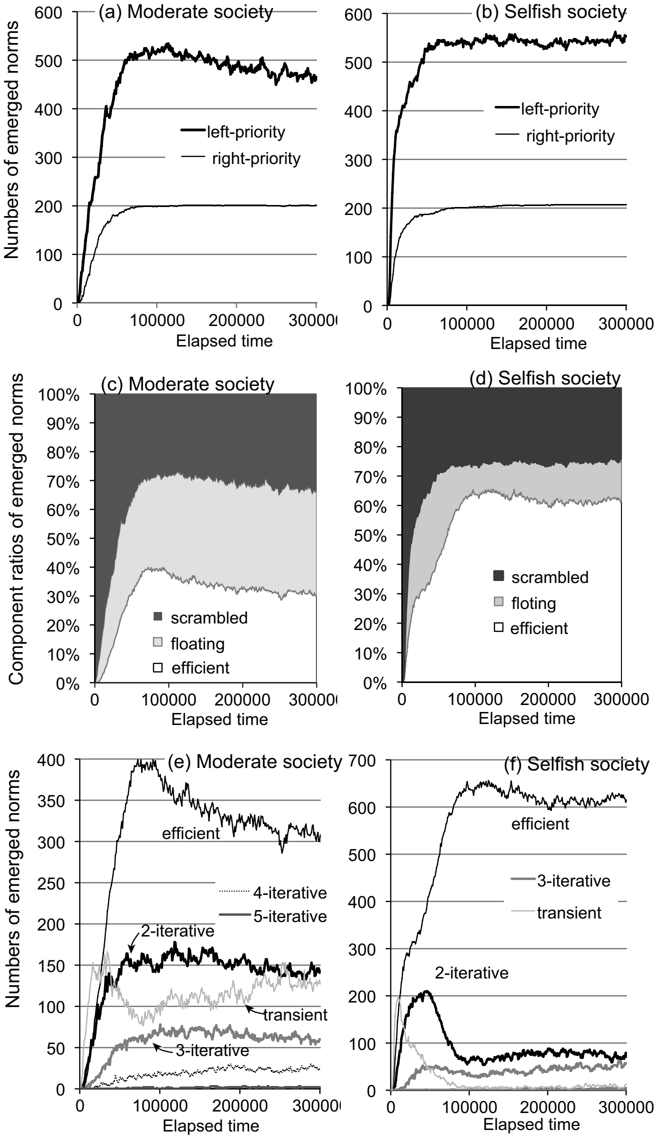

Figure 9: Ratios and distributions of emerged conventions (heterogeneous society).

4 Remarks

We are interested in the emergence of social conventions that may incur a certain small cost/penalty to a number of agents but are beneficial to the society as a whole. This kind of convention plays a significant role in conflict situations where at least one of the conflicting agents has to compromise. Furthermore, a failure to resolve the conflict means that the situation still exists, so the game recurs. Our iterative model using a non-coordinated game is a natural way to describe the conflict situation.

Another finding is that we have to separately consider the processes of convention emergence and sustainment. For example, the agents having generous matrices that have two best joint actions (p, s) and (s, p), which bring the maximal payoff (and terminate the game), can learn one of these joint actions as conventions. However, a small chance of exploration defined by ε changes the emergent efficient conventions into floating ones; that is, the chance of game iterations increases. However, if $\varepsilon = 0$, agents cannot adapt to the changes in their environments.

Similarly, a small number of self-centered agents in the moderate society can destroy the emergent efficient conventions in the adversary party by exploration and cause the scrambled conventions to increase after the conventions emerge. However, as mentioned before, the emergent conventions are relatively robust if the majority agents are selfish.

A small number of heterogeneous agents considerably change the emergent conventions. The details are not described in this paper, but we also examined a case in which A_L consisted of only moderate agents and A_R consisted of only self-centered agents. In this case, right- and left-priority conventions emerged with almost the same probability ($\mathcal{R}_R = 49.3\%$); 80% of the emergent conventions were efficient and no k-iterative conventions for $k \geq 3$ occurred. However, by replacing two agents in A_R with moderate agents, the left-priority conventions emerged more often ($\mathcal{R}_R = 23.9\%$). One future task will be is to clarify the effects of small heterogeneous agents on emergent conventions.

5 Conclusion

We discussed the question of whether conventions for resolving conflicts emerge as a result of reinforcement learning. We showed that the types of agents making up the society affect the efficiency of conflict resolution; they affect the emergent convention types and their stability. After that, we examined the features of the emergent conventions in societies consisting of unbalanced parties and those having a small number of heterogeneous agents. Our results showed that selfish agents, which have a large positive payoff for its own advantage and a small negative payoff for other's advantage, lead to efficient and sustainable Social conventions. However, they cannot devise conventions if they also have a large negative payoff for the adversary's advantage.

We plan to perform a number of experiments in different situations in the future. For example, in service computing environments, two service components that consume identical resources are requested by two or more parties of agents that have different characteristics according to the group they belong to. This corresponds to the experiments in which A_R and A_L are individually homogeneous but have different payoff matrices. Another situation is that a number of new agents that have no conventions are added to the party: we will examine how quickly these agents can acquire the conventions.

Acknowledgement This work is, in part, supported by KAKENHI (22300056).

References

[Littman, 1994] Michael L. Littman. Markov games as a framework for multi-agent reinforcement learning. In *Proc. of ICML-94*, pages 157–163. 1994.

[Moriyama and Numao, 2003] Koichi Moriyama and Masayuki Numao. Self-Evaluated Learning Agent in Multiple State Games. In *Proc. of ECML-03 (LNCS 2837)*, pages 289–300. 2003.

[Mukherjee *et al.*, 2008] Partha Mukherjee, Sandip Sen, and Stéphane Airiau. Norm emergence under constrained interactions in diverse societies. In *Proc. of AAMAS-08*, pages 779–786. 2008.

[Pujol *et al.*, 2005] Josep M. Pujol, Jordi Delgado, Ramon Sangüesa, and Andreas Flache. The role of clustering on the emergence of efficient social conventions. In *Proc. of IJCAI-05*, pages 965–970, San Francisco, CA, 2005.

[Savarimuthu *et al.*, 2008] Bastin Tony Roy Savarimuthu, Maryam Purvis, and Martin Purvis. Social norm emergence in virtual agent societies. In *Proc. of AAMAS-08*, pages 1521–1524. 2008.

[Sen and Airiau, 2007] Sandip Sen and Stéphan Airiau. Emergence of Norms Through Social Learning. In *Proc. of IJCAI-07*, pages 1507–1512, 2007.

[Shoham and Tennenholtz, 1992] Yoav Shoham and Moshe Tennenholtz. On the synthesis of useful social laws for artificial agent societies. In *Proc. of AAAI-92*, pages 276–281, 1992.

[Shoham and Tennenholtz, 1997] Yoav Shoham and Moshe Tennenholtz. On the Emergence of Social Conventions: modeling, analysis, and simulations. *Artificial Intelligence*, 94:139–166, 1997.

[Walker and Wooldridge, 1995] Adam Walker and Michael Wooldridge. Understanding the emergence of conventions in multi-agent systems. In *Proc. of 1st Int. Conf. on Multi-Agent Systems*, pages 384–389, 1995.