Vulnerability and Evolution of Cooperation in the Metanorms Game

Hitoshi Yamamoto[1] and Isamu Okada[2]

1 Faculty of Business Administration, Rissho University, Japan
2 Faculty of Business Administration, Soka University, Japan

Abstract. Some research studies have pointed out that the establishment of cooperation by metanorms as proposed by Axelrod holds only for a narrow parameter space. We closely examined the conditions under which cooperation would be established by metanorms and searched out the conditions for which cooperation could continue to exist in a stable manner. It was found that cooperation could be established by maintaining variety. It was also discovered that stability in cooperation could be robustly achieved in a relatively wide parameter space by always having a few defectors in society.

Keywords: Norms and Metanorms Game, Emergence of Order, Agent Simulation, Social Vaccine

1 Introduction

Axelrod's [1] norms game and metanorms game are well known models for maintaining order in a group. As an extension of the n-person prisoner's dilemma game, the norms game introduces the behavioral principle of non-cooperation in group participants. It was shown, however, that introducing this behavioral principle by itself could lead to non-cooperation as a dominant strategy with the norm to cooperate not being established. For this reason, Axelrod introduced metanorms, that is, the punishment of group participants who do not punish a non-cooperator, which was shown by simulation to maintain cooperation in the group. Deguchi [2] analyzed metanorms using replicator dynamics and reported that metanorms support stability in cooperation. Heckathorn [5] and Horne and Cutlip [6], moreover, conducted a psychological experiment showing that metanorms exist.

However, a number of strong criticisms of Axelrod's framework exist. Yamashita et al. [10] and Galan et al. [4] emphasized that a metanorms model featuring mutual surveillance among all members of a group leads to an upper limit in the number of group members due to cognitive limits and that a system of mutual surveillance is an unrealistic, severe restriction. In light of these criticisms, extending the metanorms game to a partial group (Prietula and Conway [9]) and limiting the study to mutual surveillance in a small world network (Newth [7]) have been proposed.

It has also been pointed out that Axelrod's findings hold only for a very limited parameter space. Oda [8] states that establishment of cooperation even with

metanorms depends on the initial probability of punishment. Galan and Izquierdo [3], meanwhile, examined Axelrod [1] by computer simulation and mathematical analysis and found that the parameter space for which metanorms could stabilize cooperation was limited.

While we also argue that Axelrod's findings have limits the same as Galan and Izquierdo, we go one step further by attempting to extract sufficient conditions for making cooperation stable.

We have discovered that cooperation can be robustly maintained by introducing into the group a small number of agents who are always behaving in a non-cooperative manner. We call this the "social vaccine" effect.

2 Norms Game and Metanorms Game

In this section, we summarize Axelrod's norms game and metanorms game and replicate them with an eye toward making extensions.

2.1 Structure of Norms Game and Metanorms Game

The norms game can be treated as an extension of the n-person prisoner's dilemma game. We consider a group of N agents. Agent i can decide to either defect or cooperate. The probability of defection is expressed as boldness Bi. If agent i defects, it gets a temptation payoff of T=3. The other (N-1) agents get a hurt payoff of H=-1. If agent i cooperates, all agents get a payoff of 0.

Up to this point, we have been describing the n-person prisoner's dilemma game. The norms game, however, gives the remaining (N-1) agents an opportunity to punish a defector. Agent j sees the defection by agent i with probability s. If the defection is not seen, nothing happens and no agent's payoff is altered. If agent j sees that agent i is defecting, agent j punishes agent i with probability Vj (vengefulness). If it turns out that agent j does punish agent i, agent i gets a payoff of P=-9 and agent j an enforcement payoff of E=-2. If no punishment occurs, no agent's payoff is altered.

The above has described the norms game. The metanorms game introduces a structure that gives agent k the opportunity to punish agent j if agent k discovers that agent j saw agent i defecting but decided to inflict no punishment. If it turns out that agent k does punish agent j, agent j gets a payoff of P=-9 and agent k a payoff of E=-2. The above structures are summarized in Fig. 1.

Each agent has a chance to defect or cooperate four times. The total payoff after four such rounds is computed and used to determine whether the agent leaves any offspring behind in the next generation. In the Axelrod model, the agent's resulting payoff is compared with the group's average payoff and standard deviation. An agent with a payoff greater than the average $+\sigma$ produces two offspring while an agent within the average $\pm\sigma$ produces one offspring. An agent with a payoff less than the average $-\sigma$ produces no offspring. Here, the number of offspring is adjusted so that population N of the group does not change between generations.

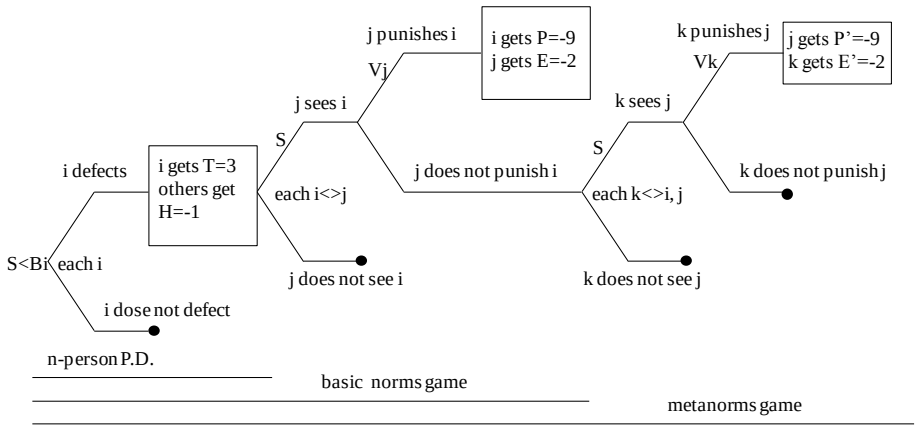

Figure 1: Structures of norms game and metanorms game (from Axelrod [1])

2.2 Axelrod Model Experiment

In this section, we replicate the norms game and metanorms game. The parameters used in this experiment are the values used by Axelrod [1].

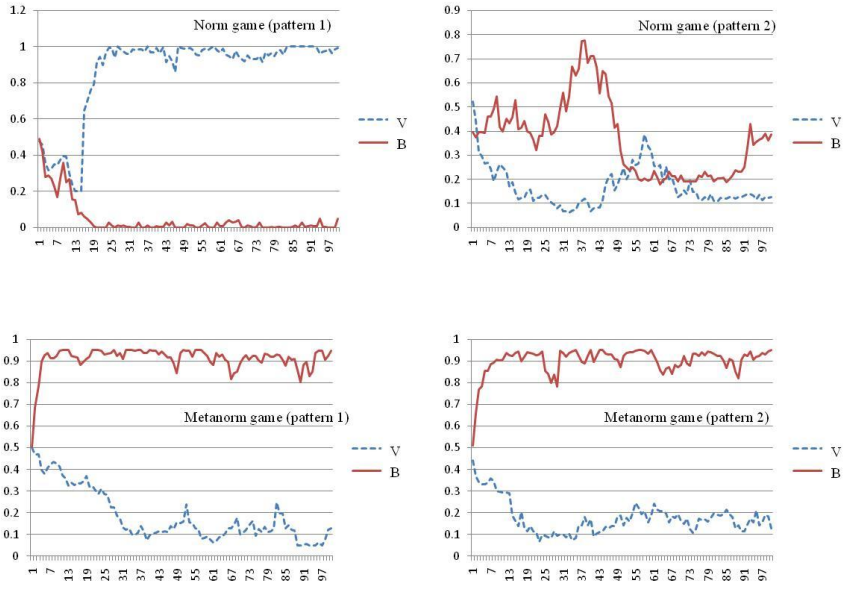

Figure 2: Behavior in the norms game and metanorms game of Axelrod's model

Table 1: Experimental parameters

Parameter	Value
Population N	20
Initial boldness	uniform random numbers [0,1]
Initial vengefulness	uniform random numbers [0,1]
Temptation payoff	T=3
Hurt payoff	H=-1
Punishment	P=-9
Enforcement payoff	E=-2
Number of generations	100
Mutation rate	0.01

Figure 2 shows the results of executing the norms game and metanorms game with different random numbers. The horizontal axis represents the number of generations and the vertical axis represents boldness (B) and vengefulness (V). As in Axelrod's experiments, the case in which the norm to cooperate is partially established with high boldness and low vengefulness otherwise (defection dominants) could be seen in the norms game, while for the metanorms game, defections were suppressed and the norm to cooperate was established.

3 Vulnerability in the Metanorms Game

In this section, we show that the metanorms game in the Axelrod model is vulnerable and search out conditions in which metanorms are established in various situations after introducing a genetic algorithm (GA).

3.1 Vulnerability in the Axelrod Model and Model Limitations

We conducted a series of experiments varying population N from 20 to 100 and number of generations from 100 to 100,000. Each experiment was executed 50 times and the average value of boldness B of the final generation was plotted (Fig. 3). Considering that vengefulness V has a strong negative correlation with B so that the behavior of V can be understood by observing the value of B, we examine only boldness B in this paper.

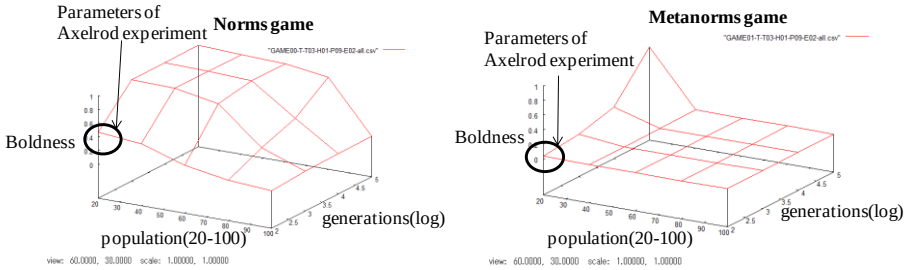

Figure 3: Axelrod model varying population and number of generations

Examining these results, it can be seen that defection tends to dominate as the number of generations increases in the norms game. This is because the intrinsic structure of the norms game makes it easy to get a free ride with respect to punishment thereby making defection dominant over the long term.

At the same time, it becomes easier to maintain cooperation as the population becomes larger. This can be explained as follows. Increasing the size of the group increases the number of times that defection can be discovered, which intensifies punishment making defection disadvantageous. This, however, implies total mutual surveillance in a large group, which is an unrealistic and severe restriction.

Turning now to the metanorms game, it can be seen that cooperation is dominant for the most part, but when increasing the number of generations at N=20, the norm to cooperate collapses. Cooperation stabilizes, however, with a slight increase in population. As previously described, total mutual surveillance is thorough at the metanorms level resulting in a very stern surveillance society that maintains cooperation.

3.2 Extension by GA Model

We have been discussing the limits of the metanorms game in various situations, but in this section, we introduce a GA into the evolutionary process. In the Axelrod model, average and standard deviation values with respect to the distribution of agent payoffs in the group are used as reference points for dividing up the group in the evolutionary process. Actual payoff distributions, however, are greatly skewed, and considering that Axelrod's method itself hints of a GA, it is only natural that a general GA be used. Figure 4 shows how similar results can be obtained for both the GA model and Axelrod model. The results shown in Fig. 4 were obtained using a GA model conducting the same experiment as that of Fig. 3.

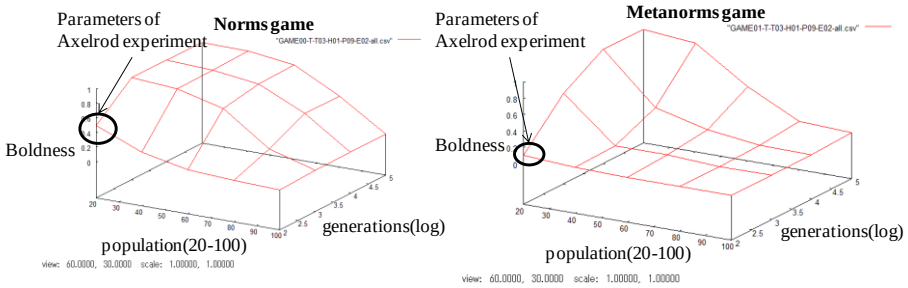

Figure 4: GA model varying population and number of generations

It can be seen from these results that collapse occurs early in the metanorms game. The reason given for this is that, while the Axelrod model depends only on mutations to achieve a new trait, adoption of the GA model makes it highly probably that new traits will be generated by crossover and that collapse will occur frequently. It can also be seen that the norm to cooperate becomes dominant as population becomes larger in either of these two models.

Continuing on, we conducted an experiment with population fixed at the basic size of N=20 while varying mutation rate. Figure 5 shows the results of varying mutation rate and the number of generations for the norms game and metanorms game. It can be seen that cooperation is established for a mutation rate of 0% and 5% or greater.

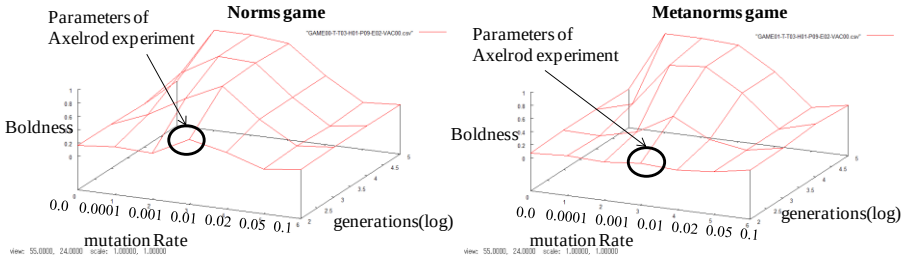

Figure 5: GA model varying mutation rate and number of generations

As described above, cooperation is established at a mutation rate of 5%, but examining results along the time-line at this mutation rate reveals a very random world.

Now, for a mutation rate of 0%, there is no primary factor for change once strategy stabilizes, and the final result is stable. In the norms game, as well, high vengefulness, which rose in value early on, promotes stability in cooperation, and since strategy becomes uniform at a low number of generations, there is no penetration of defectors and results are stable. Just before attaining uniformity, however, defection due to crossover can be seen to occur with defection becoming dominant at low frequency. As a result, average boldness becomes stable at around a value of 0.2.

4 Maintaining Cooperation by a Social Vaccine

We can summarize the results presented in the previous section as follows. For a population N=20 and a period of 100 generations, it was reconfirmed that “results of norms game = three patterns” and “metanorms = cooperation.” However, all of the three patterns (defection, intermediate, cooperation) in the norms game are part of a process toward defection since increasing the number of generations eventually results in a state in which defection is dominant. In short, the norms game converges to defection as long as the population is not increased. The metanorms game (number of agents = 20) as well results in defection over the ultra long term. This holds true even if the evolutionary process is changed to GA. Although the average defection rate is reduced at mutation rates of 0% and 5%, a mutation rate of 0% is unnatural in terms of an evolutionary game and a mutation rate of 5% results in a high degree of randomness.

In light of the above, we propose the introduction of a “social vaccine” as a policy for robustly maintaining cooperation. A vaccine, in general, refers to the inoculation of a human body with a weakened pathogen to create antibodies and ward off infection by that pathogen. A social vaccine, in turn, refers to the “inoculation” of a

group with a small number of defecting agents so as to robustly maintain norms throughout the group.

Figure 6 shows average boldness when introducing vaccine agents (agents that are always defecting) so that they make up 5% of the population in the group. These results were obtained while varying population and the number of generations. The reason for the 5% value is that the introduction of only one such agent into the smallest population in our studies (N=20) corresponds to 5% of that population.

From these results, it can be seen that defection is dominant in the norms game, but that cooperation is stable in the metanorms game even when varying the number of generations.

The reason for a metanorms collapse without a social vaccine can be given as follows. Even if agents with low vengefulness were to penetrate the group in which cooperation has been established, the fact that there is no defecting behavior means that such agents cannot be discovered with the result that agents with low vengefulness come to spread throughout the group. However, by including vaccine agents in the group, agents with low vengefulness can be easily discovered thereby preventing a drop in the group's overall level of vengefulness.

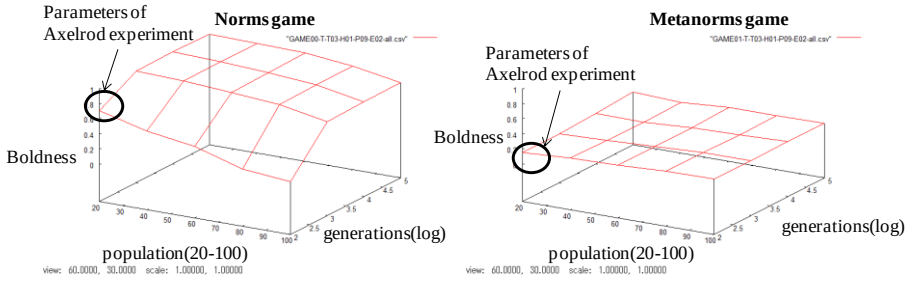

Figure 6: Stability in metanorms by introducing a social vaccine (while varying population)

Next, Fig. 7 shows experimental results when varying mutation rate and the number of generations. It can be seen that cooperation is robustly maintained in the metanorms game even for different mutation rates.

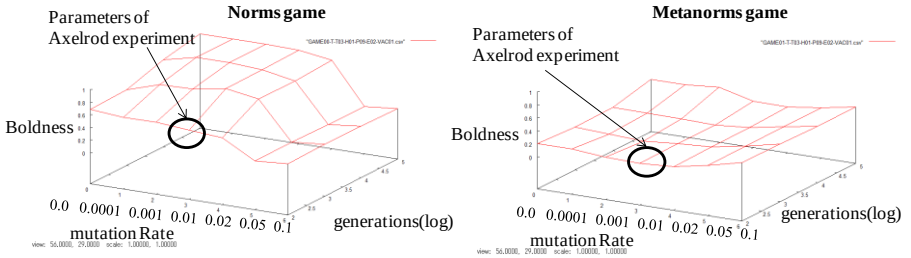

Figure 7: Stability in metanorms by introducing a social vaccine (while varying mutation rate)

5 Conclusion

Since the useful finding that metanorms are effective in maintaining stable cooperation was announced, many studies have been made taking that stability as a precondition. It has been pointed out, however, that the parameter space in which metanorms can stabilize cooperation is limited. We conducted simulations experiments to search out conditions for which metanorms can stabilize cooperation and showed that cooperation collapses in many parameter environments. Additionally, we proposed the introduction of a "social vaccine" as a policy for robustly maintaining cooperation even in parameter spaces in which cooperation would be expected to collapse. We showed that such a social vaccine can achieve ultra-long-term stability in the metanorms game for various mutation rates.

References

1) Axelrod, R.M., An Evolutionary Approach to Norms, *American Political Science Review*, 80 (4), 1095-1111, 1986.
2) Deguchi, H., Norm Game and Indirect Regulation of Multi Agent Society, *Proc. of Computational Social and Organizational Science Conference*, 2000.
3) Galan, J.M. and L.R. Izquierdo, Appearances Can Be Deceiving: Lessons Learned Re-Implementing Axelrod's 'Evolutionary Approach to Norms', *Journal of Artificial Societies and Social Simulation 8(3)*, http://jasss.soc.surrey.ac.uk/8/3/2.html, 2005.
4) Galan, J.M., M. Latek, M. Tsvetovat, and S. Rizi, Axelrod's Metanorm Games on Complex Networks, *Proc. of Agent 2007 Conference*, 271-280, 2007.
5) Heckathorn, D.D., Collective Sanctions and Compliance Norms: A Formal Theory of Group-Mediated Social Control, *American Sociological Review*, 55(3), 366-384, 1990.
6) Horne, C., and A. Cutlip, Sanctioning Costs and Norm Enforcement: An Experimental Test, *Rationality and Society,* 14(285), DOI: 10.1177/1043463102014003002, 2002
7) Newth, D., Altruistic Punishment, Social Structure and the Enforcement of Social Norms, in R. Khosla et al. (Eds.): KES 2005, LNAI 3683, 806-812, 2005.
8) Truya Oda, Evolutional Approach to the Emergence Problem of Order - application of metanorms game -, *Sociological Theory and Methods*, 5(1), 81-99, 1990.
9) Prietula, M.J. and D. Conway, The evolution of metanorms: quis custodiet ipsos custodes?, *Computational Mathematical Organization Theory*, DOI 10.1007/s10588-009-9056-4, 2009.
10) Yamashita, T., H. Kawamura, M. Yamamoto, and A. Ohuchi, Effects of Propotion of Metanorm Players on Establishment of Norm, *Fourth International Conference on Computational Intelligence and Multimedia Applications* (ICCIMA'01), 2001.