

Does Imperfect Information Always Disturb the Evolution of Reciprocity?

Shun Kurokawa^{1,2,*}

¹The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
²Kyoto University, Oiwake-cho, Kitashirakawa, Sakyo-ku, Kyoto 606-8502, Japan

*Author for correspondence (kurokawa@kais.kyoto-u.ac.jp)

Reciprocity has long been regarded as a potential explanatory mechanism for the maintenance of cooperation. However, a possible problematic case relevant to the theory of reciprocity evolution arises when the information about an opponent's behavior is imperfect. Although it has been confirmed that imperfect information disturbs the evolution of reciprocity, this argument is based on the assumption that those who attempt to cooperate always succeed in doing so. In reality, mistakes can occur, and previous studies have demonstrated that this can sway the evolution of reciprocity. In this study, removing the assumption that mistakes do not occur, we examine whether imperfect information disturbs the evolution of reciprocity in the iterated prisoner's dilemma game with errors in behavior. It might be expected that when mistakes occur, reciprocity can evolve more in the case of imperfect information than in the case of perfect information. This is because in the former case, reciprocators can miss defections incurred by other reciprocators' mistakes owing to imperfect information, which allows cooperation to persist. Contrary to this expectation, however, our analysis reveals that imperfect information still disturbs the evolution of reciprocity when mistakes occur. Additionally, we have determined that the condition under which reciprocity evolves remains unaffected, whatever reciprocators subsequently do when the opponent's last behavior was missed.

Keywords

cooperation, reciprocity, evolution, imperfect information, mistakes, optimism

Introduction

It has been a major challenge in evolutionary biology to understand the evolution of cooperation in terms of natural selection (Hamilton, 1964; Trivers, 1971). The concept of direct reciprocity has been essential for explaining why animals sometimes behave cooperatively even toward unrelated individuals (Trivers, 1971; Axelrod & Hamilton, 1981).

It has also been known that the disturbance mechanism, which involves imperfect information owing to cognitive capacity limitations (i.e., people occasionally overlook the opponent's behavior), can prevent reciprocity from evolving (Bowles & Gintis, 2011). Reciprocity is a mechanism in which cooperators respond to their opponent's behavior by switching their own behavior, thus helping the cooperation to evolve. Therefore, when information about the opponent's behavior is somehow blocked, reciprocity does not function well. In fact, it has been shown that reciprocity is less likely to evolve when information is imperfect than when information is perfect.

However, this argument on the likelihood of reciprocity under conditions of perfect and imperfect information is based on the assumption that those who attempt to cooperate always succeed in doing so. In reality, errors in behavior can occur (May, 1987). All animals, including humans, are prone to error, with attempts at cooperation sometimes ending in failure. Moreover, previous studies (May, 1987) have revealed that the existence of mistakes can dramatically sway the evolution of reciprocity because in such cases, the opponent can react to the reciprocator's mistakes as if it were a deliberate defection, starting a chain of defection that stops the current cooperation.

In this study, we consider the case in which mistakes can occur. In addition, we compare the simultaneous case in which both mistakes occur and information is imperfect with the case in which mistakes occur and information is perfect in order to examine whether imperfect information disturbs the evolution of reciprocity even when mistakes occur. It might be plausible that the evolution of reciprocity is more likely in the case of imperfect information than in the case of perfect information. We will further explain this supposition in the following sections.

In general, reciprocity is more likely to evolve if cooperation persists when two reciprocators meet each other and is more likely to evolve if cooperation stops when a reciprocator meets a defector. With this in mind, we will compare the simultaneous case in which both mistakes occur and information is imperfect with the case in which mistakes occur and information is perfect.

First, we consider the case where two reciprocators meet each other. If information is imperfect and each reciprocator sometimes or always attempts to cooperate even when there is no access to information, it is possible that even if one reciprocator mistakenly defects, the opponent reciprocator would cooperate because their imperfect information would leave the opponent reciprocator unaware of the mistake. Based on this, it seems plausible that cooperation will persist more often when information is imperfect than when information is perfect, thus contributing to a positive outcome in the evolution of reciprocity.

Second, we consider the case where a reciprocator meets a defector. When information is perfect, the reciprocator can stop cooperating with the defector.

However, when information is imperfect, the reciprocator will not know what the opponent defector does, thus causing the reciprocator to persist in cooperation and allowing the opponent to enjoy the exploitation of the reciprocator. This would lead to giving a negative outcome in the evolution of reciprocity.

Thus, it is seen that imperfect information has both positive and negative aspects for the evolution of reciprocity. It is consequently not obvious as to whether reciprocity is more likely to evolve when information is perfect or when it is imperfect, in considering the case in which mistakes can occur.

Furthermore, we raise one more question: are reciprocators more or less likely to evolve depending on whether they defect or cooperate when information is imperfect? We examine this in the following. On the one hand, when two reciprocators meet each other, reciprocators cooperate more if they cooperate when they have no information (i.e., when they are optimistic) than if they defect when they have no information (i.e., when they are pessimistic). Hence, reciprocator optimism has a positive effect on reciprocity evolution. On the other hand, when a reciprocator meets a defector, an optimistic reciprocator will be more likely to cooperate with the defector than a pessimistic reciprocator, which has an overall negative impact on reciprocity evolution. Thus, it is seen that having optimism has both positive and negative aspects for the evolution of reciprocity. It is consequently not obvious as to whether reciprocators are more or less likely to evolve depending on whether they defect or cooperate when information is imperfect.

In this article, we apply an evolutionarily stable strategy (ESS) analysis to obtain the condition under which cooperation evolves when mistakes occur. We then examine whether imperfect information makes the condition under which reciprocity evolves looser or more stringent. We also determine which reciprocators are more likely to evolve: pessimistic reciprocators who defect when they do not have information, or optimistic reciprocators who cooperate when they do not have information.

Model

Consider the iterated prisoner's dilemma game in which individuals have to choose to either cooperate or defect in each round. We assume that individuals are paired at random. The probability that the individuals interact more than t times in a given pair is w^t , where $0 < w < 1$. This assumption means that the expected number of interactions is $1 / (1 - w)$. Thus, as w increases, so does the number of interactions per pair.

An individual who cooperates will give an opponent an amount b at a personal cost of c , where $b > c > 0$. An individual who defects will give nothing. Here, we consider imperfect information. We use e , where $0 \leq e \leq 1$, to denote the probability that information is somehow blocked, i.e., that an individual cannot get access to the information about an opponent's behavior. We also consider errors in behavior (May, 1987). Similarly, we use μ , where $0 \leq \mu \leq 1$, to denote the probability that mistakes in behavior occur, i.e., that an individual who intends to cooperate fails to do so and defects.

Following earlier works (Axelrod & Hamilton, 1981),

we consider two strategies: always defect (ALLD) and tit-for-tat (TFT). ALLD defects no matter what the opponent does. If access to information about the opponent's behavior is available and the opponent cooperated in the previous round, TFT recognizes the opponent's behavior as cooperative. If access to information is available, but the opponent defected in the previous round, TFT recognizes the opponent's behavior as defective. If access to information about the opponent's behavior in the previous round is not available, TFT regards the opponent's behavior as cooperative with probability a ($0 \leq a \leq 1$). In this case, the reciprocator can be seen to become more optimistic as a increases. Moreover, in the first round, TFT attempts to cooperate with probability 1, while in the following rounds, TFT attempts to cooperate if and only if TFT recognizes the opponent's behavior as cooperative. However, TFT fails with probability μ to successfully cooperate even when TFT attempts to cooperate.

There is another type of mistake in the iterated prisoner's dilemma game: the error in perception (Axelrod & Dion, 1988). Players mistakenly regard cooperation as defection when errors in perception occur. Following two cases are the same from the viewpoint of mathematical models: one case is that TFT mistakenly regards cooperation as defection with probability e , and the other case is that TFT cannot access the information about an opponent's behavior with probability e and TFT never attempts to cooperate when information is not available (i.e., $a = 0$). Hence, we can say that our model is a generalization of the previous model about perception errors.

We define x as the expected number of total contributions by an individual playing TFT for a game in a group of two TFTs and define y as the expected number of total contributions by an individual playing TFT in a group consisting of one TFT and one ALLD. Then, we have (see supplementary file for detailed calculation),

$$x = \frac{(1 - \mu)(1 - w + wea)}{(1 - w)[1 - w(1 - e)(1 - \mu)]} \quad (1)$$

$$y = (1 - \mu) + \frac{wea(1 - \mu)}{(1 - w)} \quad (2)$$

Using these expected numbers, the respective payoffs ($b - c$) x accumulated through a game are described as follows: TFT gets payoff by in a game against TFT; and ALLD gets payoff in a game against TFT.

We can then determine the condition under which TFT is a strict ESS against an invasion of ALLD. The condition is that TFT's payoff against itself is larger than TFT's payoff against ALLD, given as $(b - c)x > by$. Using (1) and (2), this inequality becomes

$$\frac{c}{b} < w(1 - e)(1 - \mu) \quad (3)$$

When $w(1 - \mu) > 0$ is satisfied, the right-hand side of (3) decreases as e increases regardless of the value of μ . This indicates that, even when considering errors in behavior, imperfect information will still disturb the evolution of cooperation. The condition under which reciprocity evolves is more stringent when information is imperfect

than when it is perfect. When $w(1 - \mu) = 0$, the right-hand side of (3) is always equal to zero. Considering the assumption that $b, c > 0$, however, the inequality (3) does not hold and thus the reciprocity never evolves.

It is also apparent that the right-hand side of (3) does not contain the parameter a , indicating that the condition under which reciprocity evolves is not affected by whether the reciprocators are pessimistic or optimistic. Even though the reciprocators must choose cooperation or defection without information when information is blocked, the choice of cooperation or defection does not matter.

Discussion

In this analysis, we obtained two results: one is that imperfect information disturbs the evolution of reciprocity even when there are errors in behavior; the other result is that the condition under which reciprocity evolves is not affected by whether a reciprocator cooperates or defects when there is no access to information.

There is another type of strategy in the iterated prisoner's dilemma game called suspicious TFT, which does not cooperate in the first round and then follows the strategy of TFT after the first round (Boyd & Lorberbaum, 1987). A previous study (To, 1988), via analyses of suspicious TFT, has revealed that those who cooperate in the first round outperform those who do not. However, from this study, we cannot determine which is beneficial for the evolution of cooperation, cooperating in the first round or cooperating when information is not available, since information is not available in the first round. On the other hand, our study compares a specific TFT strategy that cooperates when information is not available, excluding the first round, with another specific TFT strategy that defects when information is not available, excluding the first round. In the result, our study has revealed that cooperating when information is not available (i.e., being optimistic) does not influence the likelihood of reciprocity evolution. Combining the data of the current study with that of the previous study, we can see that cooperating when information is not available is not beneficial for evolution; however, cooperating in the first round is beneficial for the evolution of cooperation. Behaving cooperatively when information is not available is not important. Behaving cooperatively in the first meeting is important.

This study also analyzed pairwise games to demonstrate that reciprocity is equally likely to evolve whether reciprocators are pessimistic or optimistic. It is of interest to extend this analysis to n-player games (Boyd & Richerson, 1988; Kurokawa & Ihara, 2009, 2013; Kurokawa, Wakano, & Ihara, 2010; Deng, Li, Kurokawa, & Chu, 2012), and the corresponding literature (S Kurokawa, Y Ihara, unpublished data) shows that in many-individual interactions, reciprocity is more likely to evolve when reciprocators are optimistic than when they are pessimistic. Thus, scale (in terms of the number of interacting individuals) has an effect on the evolution of optimism in theory. Any empirical studies that could show that animals (including humans) interacting in n-individual games are more optimistic than animals operating only in a pair-wise manner could be seen as an interesting confirmation of these theoretical studies.

Acknowledgement

The author thanks Tatsuya Sasaki for giving comments.

References

- Axelrod, R., & Dion, D. (1988). The further evolution of cooperation. *Science*, 242, 1385–1390. (doi: 10.1126/science.242.4884.1385)
- Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, 211, 1390–1396. (doi: 10.1126/science.7466396)
- Bowles, S., & Gintis, H. (2011). *A cooperative species: Human reciprocity and its evolution*. Princeton, NJ: Princeton University Press. (doi: 10.1515/9781400838837)
- Boyd, R., & Lorberbaum, J. P. (1987). No pure strategy is evolutionarily stable in the repeated Prisoner's Dilemma game. *Nature*, 327, 58–59. (doi: 10.1038/327058a0)
- Boyd, R., & Richerson, P. J. (1988). The evolution of reciprocity in sizable groups. *Journal of Theoretical Biology*, 132, 337–356. (doi: 10.1016/S0022-5193(88)80219-4)
- Deng, K., Li, Z., Kurokawa, S., & Chu, T. (2012). Rare but severe concerted punishment that favors cooperation. *Theoretical Population Biology*, 81, 284–291. (doi: 10.1016/j.tpb.2012.02.005)
- Hamilton, W. D. (1964). The genetical evolution of social behaviour I, II. *Journal of Theoretical Biology*, 7, 1–52. (doi: 10.1016/0022-5193(64)90038-4) (doi: 10.1016/0022-5193(64)90039-6)
- Kurokawa, S., & Ihara, Y. (2009). Emergence of cooperation in public goods games. *Proceedings of the Royal Society B*, 276, 1379–1384. (doi: 10.1098/rspb.2008.1546)
- Kurokawa, S., & Ihara, Y. (2013). Evolution of social behavior in finite populations: A payoff transformation in general n-player games and its implications. *Theoretical Population Biology*, 84, 1–8. (doi: 10.1016/j.tpb.2012.11.004)
- Kurokawa, S., Wakano, J. Y., & Ihara, Y. (2010). Generous cooperators can outperform non-generous cooperators when replacing a population of defectors. *Theoretical Population Biology*, 77, 257–262. (doi: 10.1016/j.tpb.2010.03.002)
- May, R. M. (1987). More evolution of cooperation. *Nature*, 327, 15–17. (doi: 10.1038/327015a0)
- To, T. (1988). More realism in the prisoner's dilemma. *Journal of Conflict Resolution*, 32, 402–408. (doi: 10.1177/0022002788032002008)
- Trivers, R. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46, 35–57. (doi: 10.1086/406755)