**Does imperfect information always disturb the evolution of reciprocity?**

Shun Kurokawa1,2,\*

1The University of Tokyo, Hongo 7-3-1, Bunkyoku, Tokyo 113-0033, Japan

2Kyoto University, Oiwake-cho, Kitashirakawa, Sakyo-ku, Kyoto 606-8502, Japan

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

Supplementary file

Proof for (1)

Here, we let *A* denote the expected amount of cooperation by a given TFT when both TFTs cooperate in a specific round and in all rounds thereafter. Similarly, we let *B* denote the expected amount of cooperation by a TFT in a given round and thereafter when one TFT cooperates and the other TFT defects and *C* denote the expected amount of cooperation by a TFT in a given round and thereafter when both TFTs defect.

Consider the first round. In the first round, TFT attempts to cooperate with probability 1, although failing to successfully do so with probability$ μ$. Hence, the probability of cooperating is $1-μ$. Therefore, the probability that both players cooperate is $\left(1-μ\right)^{2}$, and in that case, the expected number of contributions in a specific round and in all rounds thereafter is *A*. The probability that one player cooperates and the other player defects is $ 2μ\left(1 - μ\right)$, and in that case, the expected number of contributions in a specific round and in all rounds thereafter is *B*. The probability that both players defect is $μ^{2}$, and in that case, the expected number of contributions in a specific round and in all rounds thereafter is *C*. Hence,

$x=A\left(1-μ\right)^{2} + 2μ\left(1 - μ\right)B + Cμ^{2}$ (S1)

Consider the situation wherein both players (TFTs) cooperate in a specific round. The probability of cooperating in the following round is $\left(1-e\right)\left(1-μ\right)＋ea\left(1-μ\right)$, since TFT cooperates when TFT can get access to the information about an opponent’s behavior and succeed in cooperating without failing to do so, or when TFT cannot get access to the information about an opponent’s behavior and attempts to cooperate with probability *a* and succeeds in cooperating without failing to do so. Therefore, in the following round, the probability that both players cooperate is $\left[\left(1-e\right)\left(1-μ\right)＋ea\left(1-μ\right)\right]^{2}$, and in that case, the expected number of contributions in a specific round and in all rounds thereafter is *A*. The probability that one player cooperates and the other player defects is $ 2\left[\left(1-e\right)\left(1-μ\right)＋ea\left(1-μ\right)\right]\left\{1-\left[\left(1-e\right)\left(1-μ\right)＋ea\left(1-μ\right)\right]\right\}$, and in that case, the expected number of contributions in a specific round and in all rounds thereafter is *B*. The probability that both players defect is $[1-\{\left(1-e\right)\left(1-μ\right)＋ea\left(1-μ\right)\}]^{2}$, and in that case, the expected number of contributions in a specific round and in all rounds thereafter is *C*. Hence,

$A=1+w\{A\left[\left(1-e\right)\left(1-μ\right)＋ea\left(1-μ\right)\right]^{2}+2B\left[\left(1-e\right)\left(1-μ\right)＋ea\left(1-μ\right)\right]\left\{1-\left[\left(1-e\right)\left(1-μ\right)＋ea\left(1-μ\right)\right]\right\}+C[1-\{\left(1-e\right)\left(1-μ\right)＋ea\left(1-μ\right)\}]^{2}\}$ (S2)

Consider the situation wherein one player (TFT) cooperates and the other player (TFT) defects in a specific round. The probability that the player who cooperated in the previous round cooperates in the specific round is $ea\left(1-μ\right)$, since TFT cooperates when TFT cannot get access to the information about an opponent’s defection and attempts to cooperate with probability *a* and succeeds in cooperating without failing to do so. On the other hand, the probability that the player who defected in the previous round cooperates in the specific round is $\left(1-e\right)\left(1-μ\right)＋ea\left(1-μ\right)$, since TFT cooperates when TFT can get access to the information about an opponent’s cooperation and succeed in cooperating without failing to do so, or when TFT cannot get access to the information about an opponent’s behavior and attempts to cooperate with probability *a* and succeeds in cooperating without failing to do so. Therefore, in the following round, the probability that both players cooperate is $ea\left(1-μ\right)\left\{\left(1-e\right)\left(1-μ\right)+ea\left(1-μ\right)\right\}$, and in that case, the expected number of contributions in a specific round and in all rounds thereafter is *A*. The probability that one player cooperates and the other player defects is $\left[1-ea\left(1-μ\right)\right]\left\{\left(1-e\right)\left(1-μ\right)+ea\left(1-μ\right)\right\}+ea\left(1-μ\right)[1-\left\{\left(1-e\right)\left(1-μ\right)+ea\left(1-μ\right)\right\}]$, and in that case, the expected number of contributions in a specific round and in all rounds thereafter is *B*. The probability that both players defect is $\left[1-ea\left(1-μ\right)\right][1-\left\{\left(1-e\right)\left(1-μ\right)+ea\left(1-μ\right)\right\}]$, and in that case, the expected number of contributions in a specific round and in all rounds thereafter is *C*. Hence,

$B=1/2+w\{Aea\left(1-μ\right)\left\{\left(1-e\right)\left(1-μ\right)+ea\left(1-μ\right)\right\}+B\left\{\left[1-ea\left(1-μ\right)\right]\left\{\left(1-e\right)\left(1-μ\right)+ea\left(1-μ\right)\right\}+ea\left(1-μ\right)\left[1-\left\{\left(1-e\right)\left(1-μ\right)+ea\left(1-μ\right)\right\}\right]\right\}+C\left[1-ea\left(1-μ\right)\right]\left[1-\left\{\left(1-e\right)\left(1-μ\right)+ea\left(1-μ\right)\right\}\right]\}$ (S3)

Consider the situation wherein both players (TFTs) defect in a specific round. The probability of cooperating in the following round is $ea\left(1-μ\right)$, since TFT cooperates when TFT cannot get access to the information about an opponent’s defection and attempts to cooperate with probability *a* and succeeds in cooperating without failing to do so. Therefore, in the following round, the probability that both players cooperate is $[ea(1-μ)]^{2}$, and in that case, the expected number of contributions in a specific round and in all rounds thereafter is *A*. The probability that one player cooperates and the other player defects is $2[1-ea\left(1-μ\right)]ea\left(1-μ\right)$, and in that case, the expected number of contributions in a specific round and in all rounds thereafter is *B*. The probability that both players defect is $[1-ea(1-μ)]^{2}$, and in that case, the expected number of contributions in a specific round and in all rounds thereafter is *C*. Hence,

$C=w\{A[ea(1-μ)]^{2}+2B[1-ea\left(1-μ\right)]ea\left(1-μ\right)+C[1-ea(1-μ)]^{2}\}$ (S4)

Solving (S2)–(S4), we obtain

$A=\frac{[1-w\left\{1-ea\left(1-μ\right)\right\}]}{\left(1-w\right)[1-w\left(1-e\right)\left(1-μ\right)]}$ (S5)

$B=\frac{[1+w\left\{2ea\left(1-μ\right)-1\right\}]}{2(1-w)[1-w\left(1-e\right)\left(1-μ\right)]}$ (S6)

$C=\frac{wea\left(1-μ\right)}{(1-w)[1-w\left(1-e\right)\left(1-μ\right)]}$ (S7)

Substituting (S5)–(S7) into (S1) gives

$x=\frac{(1-μ)(1-w+wea)}{(1-w)[1-w(1-e)\left(1-μ\right)]}$

This is end of the proof.

Proof for (2)

In the first round, TFT attempts to cooperate with probability 1, although failing to do so successfully with probability$ μ$. Hence, the probability of cooperating is $1-μ$. In the following rounds, TFT cooperates when and only when TFT cannot access the information about an opponent (ALLD)’s defection and attempts to cooperate with probability *a* and succeeds in cooperating without failing to do so. Hence, the probability of cooperating is $ea\left(1-μ\right)$. Therefore,

$y=\left(1-μ\right)+ea\left(1-μ\right)w/(1-w)$

This is end of the proof.