Supplementary Information

Introspection dynamics in asymmetric multiplayers games

Marta Couto[1] and Saptarshi Pal[1]

[1]Max Planck Research Group Dynamics of Social Behavior, Max Planck Institute for Evolutionary Biology, 24306 Ploen, Germany

Abstract

Introduction

Model

Appendix: Proofs

Proof. **Proof of Proposition ??**

Since β is finite, the transition matrix of the process $\mathbf{T}$ given by Eq. **??** is primitive and therefore, the stationary distribution $\mathbf{u} = (\mathbf{u_a})_{\mathbf{a}\in\mathbf{A}}$ of the row-stochastic transition matrix is unique and satisfies the conditions laid out in Eq. **??** and **??**. To continue for the rest of the proof, we introduce some short-cut notation that will be of use later in the proof:

$$I_q := I(\mathbf{a}_q, \mathbf{a}), \quad \textit{iff} \quad \mathbf{a}_q \in Neb(\mathbf{a}) \tag{1}$$

$$\tau^j_{\mathbf{a}^j} := \frac{1}{\sum\limits_{a' \in \mathbf{A}^j} e^{\beta\left(\pi^j_{a'} - \pi^j_{\mathbf{a}^j}\right)}} \tag{2}$$

In order to show that the candidate stationary distribution, as proposed in Eq. **??** is the stationary distribution of the process, we need to show that the following are true:

$$\mathbf{T}_{\mathbf{a},\mathbf{a}}\mathbf{u_a} + \sum_{\mathbf{a}_q \neq \mathbf{a}} \mathbf{T}_{\mathbf{a}_q,\mathbf{a}}\mathbf{u}_{\mathbf{a}_q} = \mathbf{u}_a \quad \forall \mathbf{a} \in \mathbf{A} \tag{3}$$

$$\sum_{\mathbf{a}\in\mathbf{A}} \mathbf{u_a} = 1 \tag{4}$$

The Eq 3 can be simplified further with the steps:

$$\mathbf{T}_{\mathbf{a},\mathbf{a}}\mathbf{u_a} + \sum_{\mathbf{a}_q \neq \mathbf{a}} \mathbf{T}_{\mathbf{a}_q,\mathbf{a}}\mathbf{u}_{\mathbf{a}_q} = \tag{5}$$

$$\left(1 - \frac{1}{N} \sum_{\mathbf{a}_q \in Neb(\mathbf{a})} \frac{1}{m^{I_q} - 1} \cdot p^{I_q}_{\mathbf{a}^{I_q} \to \mathbf{a}_q^{I_q}} \cdot \mathbf{u_a}\right) + \sum_{\mathbf{a}_q \in Neb(\mathbf{a})} \frac{1}{m^{I_q} - 1} \cdot p^{I_q}_{\mathbf{a}_q^{I_q} \to \mathbf{a}^{I_q}} \cdot \mathbf{u}_{\mathbf{a}_q} = \tag{6}$$

$$\mathbf{u_a} + \frac{1}{N} \sum_{\mathbf{a}_q \in Neb(\mathbf{a})} \left(\prod_{k \neq I_q} \tau^k_{\mathbf{a}^k}\right) \left(p^{I_q}_{\mathbf{a}_q^{I_q} \to \mathbf{a}^{I_q}} \cdot \tau^{I_q}_{\mathbf{a}^{I_q}} - p^{I_q}_{\mathbf{a}^{I_q} \to \mathbf{a}_q^{I_q}} \cdot \tau^{I_q}_{\mathbf{a}_q^{I_q}}\right) \cdot \left(\frac{1}{m^{I_q} - 1}\right) \tag{7}$$

Using the definition of probability of update of the introspection dynamics, as given by Eq. **??** and Eq. 2, it can be shown that:

$$p^{I_q}_{\mathbf{a}_q^{I_q} \to \mathbf{a}^{I_q}} \cdot \tau^{I_q}_{\mathbf{a}^{I_q}} - p^{I_q}_{\mathbf{a}^{I_q} \to \mathbf{a}_q^{I_q}} \cdot \tau^{I_q}_{\mathbf{a}_q^{I_q}} = 0 \tag{8}$$

Plugging the equality in Eq. 8 into Eq. 7, we can see that the left hand side of Eq. 3 indeed simplifies to $\mathbf{u_a}$. Now, to confirm that the candidate $\mathbf{u}$ is the unique stationary distribution we need to check if Eq. 4 holds. Simplifying the left hand side of this equation shows that:

$$\sum_{\mathbf{a}\in\mathbf{A}} \mathbf{u_a} = \sum_{\mathbf{a}\in\mathbf{A}} \prod_{k=1}^{N} \tau^k_{\mathbf{a}^k} \tag{9}$$

$$= \left(\prod_{k=1}^{N} \sum_{\mathbf{a}'\in\mathbf{A}} e^{\beta\pi^k_{\mathbf{a}'}}\right)^{-1} \cdot \left(\sum_{\mathbf{a}\in\mathbf{A}} \prod_{k=1}^{N} e^{\beta\pi^k_{\mathbf{a}^k}}\right) \tag{10}$$

$$= 1 \tag{11}$$

The step from Eq. 10 to Eq. 11 is possible because the sum and product in Eq. 10 are interchangeable for both the terms. Therefore, condition Eq. 4 is satisfied too. □

Proof. **Proof of Proposition ??**

If $\mathbf{u}$ is the unique stationary distribution of the $N-$player additive game with under the finite selection introspection dynamics, it is given by the expression in Eq. **??**. We calculate the marginal distribution of any arbitrary state $\mathbf{a}$, $\xi_{\mathbf{a}} = (\xi^j_{\mathbf{a}^j})_{j=1,2,\ldots,N}$ by using the definition of marginal distribution in Eq. **??**. It follows that:

$$\xi^j_{\mathbf{a}^j} = \sum_{\mathbf{b}\in\mathbf{A}^{-j}} \mathbf{u}_{(\mathbf{a}^j,\mathbf{b})} \tag{12}$$

$$= \left(\prod_{k=1}^{N} \sum_{\mathbf{a}'\in\mathbf{A}^k} e^{\beta\pi^k_{\mathbf{a}'}}\right)^{-1} \cdot e^{\beta\pi^j_{\mathbf{a}^j}} \cdot \left(\sum_{\mathbf{b}\in\mathbf{A}^{-j}} \prod_{k\neq j} e^{\beta\pi^k_{\mathbf{b}^k}}\right) \tag{13}$$

$$= \left(\sum_{\mathbf{a}'\in\mathbf{A}^j} e^{\beta\pi^j_{\mathbf{a}'}}\right)^{-1} \cdot e^{\beta\pi^j_{\mathbf{a}^j}} \cdot \left(\prod_{k\neq j} \sum_{\mathbf{a}'\in\mathbf{A}^k} e^{\beta\pi^k_{\mathbf{a}'}}\right)^{-1} \cdot \left(\sum_{\mathbf{b}\in\mathbf{A}^{-j}} \prod_{k\neq j} e^{\beta\pi^k_{\mathbf{b}^k}}\right) \tag{14}$$

$$= \left(\sum_{\mathbf{a}'\in\mathbf{A}^j} e^{\beta\pi^j_{\mathbf{a}'}}\right)^{-1} \cdot e^{\beta\pi^j_{\mathbf{a}^j}} \cdot \left(\sum_{\mathbf{a}'\in\mathbf{A}^{-j}} \prod_{k\neq j} e^{\beta\pi^k_{\mathbf{a}'}}\right)^{-1} \cdot \left(\sum_{\mathbf{b}\in\mathbf{A}^{-j}} \prod_{k\neq j} e^{\beta\pi^k_{\mathbf{b}^k}}\right) \tag{15}$$

$$= \left(\sum_{\mathbf{a}'\in\mathbf{A}^j} e^{\beta\left(\pi^j_{\mathbf{a}'}-\pi^j_{\mathbf{a}^j}\right)}\right)^{-1} \tag{16}$$

Therefore, the marginal distribution follows Eq. **??**. Now since we also additionally know that the stationary distribution follows the form Eq. **??**, we can conclude that for additive games, under introspection dynamics with finite selection, Eq. **??** holds. □

Proof. **Proof of Proposition ??**

Since we have demonstrated that the linear public goods game is an additive game, the proof of this theorem can be performed by directly using Theorem **??**. Here, we provide an independent proof. The idea behind this proof is identical to the proof of Theorem **??**.

The stationary transition matrix $\mathbf{T}$ for the linear public goods game is primitive when β is finite (i.e., there is a positive power k such that $\mathbf{T}^k$ is a strictly positive matrix). Therefore, the stationary distribution of $\mathbf{T}$ will always be unique. We define the following short cut notations for the ease of the proof:

$$\bar{\mathbf{a}}^j := \{\mathrm{D},\mathrm{C}\} \setminus \mathbf{a}^j \tag{17}$$

$$p^j := \frac{1}{1+e^{\beta f(c^j,r^j)}} \tag{18}$$

In addition we introduce a mapping function $\alpha(.)$ which maps the action C to 1 and the action D to 0.

That is $\alpha(\mathrm{C}) := 1$ and $\alpha(\mathrm{D}) := 0$. Using these notations and Eq. **??** and **??** we can write the probability that a player j updates from $\mathbf{a}^j$ to $\bar{\mathbf{a}}^j$ while their co-players play $\mathbf{a}^{-j}$ as:

$$p^j_{\bar{\mathbf{a}}^j \to \mathbf{a}^j}(\mathbf{a}^{-j}) = p^j\, sign(\mathbf{a}^j) + \alpha(\bar{\mathbf{a}}^j) \tag{19}$$

The candidate stationary distribution $\mathbf{u}$ given in Eq **??** can be written down using our shortcut notation as:

$$\mathbf{u}_{\mathbf{a}} = \prod_{k=1}^{N} p^k\, sign(\mathbf{a}^k) + \alpha(\bar{\mathbf{a}}^k) \quad , \forall \mathbf{a} \in \{0,1\}^N \tag{20}$$

This stationary distribution must satisfy the following properties, which are also given in Eq **??** and **??**:

$$\mathbf{u}_{\mathbf{a}} = \mathbf{T}_{\mathbf{a},\mathbf{a}} \mathbf{u}_{\mathbf{a}} + \sum_{\mathbf{a}_q \neq \mathbf{a}} \mathbf{T}_{\mathbf{a}_q,\mathbf{a}} \mathbf{u}_{\mathbf{a}_q} \tag{21}$$

$$\sum_{\forall \mathbf{a}_q} \mathbf{u}_{\mathbf{a}_q} = 1 \tag{22}$$

Where, the terms in the right hand side of Eq. 21 can be simplified using Eq **??** and **??** as follows:

$$\mathbf{T}_{\mathbf{a},\mathbf{a}} = 1 - \sum_{k=1}^{N} \mathbf{T}_{(\mathbf{a}^k,\mathbf{a}^{-k}),(\bar{\mathbf{a}}^k,\mathbf{a}^{-k})} = 1 - \frac{1}{N} \sum_{k=1}^{N} p^k sign(\bar{\mathbf{a}}^k) + \alpha(\mathbf{a}^k) \tag{23}$$

and additionally, also using Eq. 20 the second term can be simplified too:

$$\sum_{\mathbf{a}_q \neq \mathbf{a}} \mathbf{T}_{\mathbf{a}_q,\mathbf{a}} \mathbf{u}_{\mathbf{a}_q} = \sum_{k=1}^{N} \mathbf{T}_{(\bar{\mathbf{a}}^k,\mathbf{a}^{-k}),(\mathbf{a}^k,\mathbf{a}^{-k})} \mathbf{u}_{(\bar{\mathbf{a}}^k,\mathbf{a}^{-k})} \tag{24}$$

$$= \frac{1}{N} \sum_{k=1}^{N} \left(p^k sign(\mathbf{a}^k) + \alpha(\bar{\mathbf{a}}^k) \right) \mathbf{u}_{(\bar{\mathbf{a}}^k,\mathbf{a}^{-k})} \tag{25}$$

$$= \frac{\mathbf{u}_{\mathbf{a}}}{N} \sum_{k=1}^{N} p^k sign(\bar{\mathbf{a}}^k) + \alpha(\mathbf{a}^k) \tag{26}$$

Now, using Eq. 23, 26 one can show that the right hand side of Eq. 21 is the element of the stationary distribution, corresponding to the state $\mathbf{a}$, $\mathbf{u}_a$. Now, to complete the proof, we must show that Eq. 22 is also true for our candidate stationary distribution. This can be done by decomposing the sum of the elements of the stationary distribution as follows:

$$\sum_{\forall \mathbf{a}_q} u_{\mathbf{a}_q} = \sum_{\forall \mathbf{a}_q} \prod_{k=1}^{N} p^k sign(\mathbf{a}_q^k) + \alpha(\bar{\mathbf{a}}_q^k) \tag{27}$$

$$= \sum_{\forall \mathbf{a}_q} (1 - p^N) \prod_{k=1}^{N-1} p^k sign(\mathbf{a}_q^k) + \alpha(\bar{\mathbf{a}}_q^k) + p^N \prod_{k=1}^{N-1} p^k sign(\mathbf{a}_q^k) + \alpha(\bar{\mathbf{a}}_q^k) \tag{28}$$

$$= \sum_{\forall \mathbf{a}_q} \prod_{k=1}^{N-1} p^k sign(\mathbf{a}_q^k) + \alpha(\bar{\mathbf{a}}_q^k) \tag{29}$$

When the above decomposition is perfomed $N-1$ more times, the sum of the right hand side becomes 1. This prooves that the candidate stationary distribution is also a probability distribution. □

Proof. **Proof of Proposition ??**

By construction, the candidate stationary distribution given by Eq. **??** and Eq. **??** is a probability distribution since it satisfies the condition in Eq. **??** and for any state $\mathbf{a}'$, $\mathbf{u}_{\mathbf{a}'}$ is between 0 and 1. Moreover, since β is finite, the transition matrix of the process $\mathbf{T}$ is primitive and therefore, it will have a unique stationary distribution. To show that the candidate stationary distribution is the unique stationary distribution, we need to check if for this process, $\mathbf{uT} = \mathbf{u}$. That is, the condition in Eq. 21 must hold for all states $\mathbf{a}$. We re-introduce some notations that we will use in this proof:

$$\bar{\mathbf{a}}^j := \{\mathrm{D}, \mathrm{C}\} \setminus \mathbf{a}^j \tag{30}$$

$$\alpha(a) := \begin{cases} 1 & \text{if} \quad a = \mathrm{C} \\ \\ 0 & \text{if} \quad a = \mathrm{D} \end{cases} \tag{31}$$

$$\mathcal{C}(\mathbf{a}) = \sum_{j=1}^{N} \alpha(\mathbf{a}^j) \tag{32}$$

For this process, the first term in the right hand side of Eq. 21 can be simplified as:

$$\mathbf{u}_{\mathbf{a}} \mathbf{T}_{\mathbf{a},\mathbf{a}} = \mathbf{u}_{\mathbf{a}} - \mathbf{u}_{\mathbf{a}} \sum_{k=1}^{N} \mathbf{T}_{(\mathbf{a}^k, \mathbf{a}^{-k}), (\bar{\mathbf{a}}^k, \mathbf{a}^{-k})} \tag{33}$$

$$= \mathbf{u}_{\mathbf{a}} - \frac{\mathbf{u}_{\mathbf{a}}}{N} \sum_{k=1}^{N} \frac{1}{1 + e^{sign(\bar{\mathbf{a}}^k) \beta f(N_k)}} \tag{34}$$

Where, the function $sign(.)$ is defined as in Eq. **??** and $f(j)$ is the difference in payoffs between playing D and C when there are j co-players playing C. The term N_k is the number of co-players of k that play C in state $\mathbf{a}$. That is,

$$N_k := \sum_{j \neq k} \alpha(\mathbf{a}^j) \tag{35}$$

The second term in the right hand side of Eq. 21 can be simplified as,

$$\sum_{\mathbf{a}_q \neq \mathbf{a}} \mathbf{T}_{\mathbf{a}_q,\mathbf{a}} \mathbf{u}_{\mathbf{a}_q} = \sum_{k=1}^{N} \mathbf{T}_{(\bar{\mathbf{a}}^k,\mathbf{a}^{-k}),(\mathbf{a}^k,\mathbf{a}^{-k})} \mathbf{u}_{(\bar{\mathbf{a}}^k,\mathbf{a}^{-k})} \tag{36}$$

$$= \frac{1}{N\Gamma} \sum_{k=1}^{N} \mathbf{T}_{(\bar{\mathbf{a}}^k,\mathbf{a}^{-k}),(\mathbf{a}^k,\mathbf{a}^{-k})} \prod_{j=1}^{\mathcal{C}((\bar{\mathbf{a}}^k,\mathbf{a}^{-k}))} e^{-\beta f(j-1)} \tag{37}$$

$$= \frac{1}{N\Gamma} \sum_{k=1}^{N} \mathbf{T}_{(\bar{\mathbf{a}}^k,\mathbf{a}^{-k}),(\mathbf{a}^k,\mathbf{a}^{-k})} \left(\prod_{j=1}^{N_k} e^{-\beta f(j-1)} \right) \cdot e^{-\beta\alpha(\bar{\mathbf{a}}^k) f(-\alpha(\mathbf{a}^k)+N_k} \tag{38}$$

Between the steps, Eq. 37 and 38, we took out one term from the product that is present in our candidate distribution. This term accounts for the k^{th} players action in the neighbouring state $(\bar{\mathbf{a}}^k, \mathbf{a}^{-k})$ of $\mathbf{a}$. For simplicity, we replace $\mathbf{T}_{(\bar{\mathbf{a}}^k,\mathbf{a}^{-k}),(\mathbf{a}^k,\mathbf{a}^{-k})}$ with just $\mathbf{T}$ in the next steps. We continue the simplification of Eq. 38 in the later steps by introducing terms that cancel each other.

$$\sum_{\mathbf{a}_q \neq \mathbf{a}} \mathbf{T}_{\mathbf{a}_q,\mathbf{a}} \mathbf{u}_{\mathbf{a}_q} = \frac{1}{N\Gamma} \sum_{k=1}^{N} \mathbf{T} \cdot \left(\prod_{j=1}^{N_k} e^{-\beta f(j-1)} \right) \cdot \frac{e^{-\beta\alpha(\bar{\mathbf{a}}^k) f(-\alpha(\mathbf{a}^k)+N_k)}}{e^{-\beta\alpha(\mathbf{a}^k) f(-\alpha(\bar{\mathbf{a}}^k)+N_k)}} \cdot e^{-\beta\alpha(\mathbf{a}^k) f(-\alpha(\bar{\mathbf{a}}^k)+N_k)} \tag{39}$$

The newly introduced term in Eq. 39 can be taken inside the product. Note that this term is 0 if the k^{th} player plays D in the state $\mathbf{a}$. When this term is taken inside the product bracket, products of exponent $e^{-\beta f(j-1)}$ can be performed for j ranging from 1 to the number of cooperators in state $\mathbf{a}$, $\mathcal{C}(\mathbf{a})$. This product is then the stationary distribution probability $\mathbf{u_a}$. That is,

$$\sum_{\mathbf{a}_q \neq \mathbf{a}} \mathbf{T}_{\mathbf{a}_q,\mathbf{a}} \mathbf{u}_{\mathbf{a}_q} = \frac{1}{N\Gamma} \sum_{k=1}^{N} \mathbf{T} \cdot \left(\prod_{j=1}^{N_k} e^{-\beta f(j-1)} \cdot e^{-\beta\alpha(\mathbf{a}^k) f(-\alpha(\bar{\mathbf{a}}^k)+N_k)} \right) \cdot \frac{e^{-\beta\alpha(\bar{\mathbf{a}}^k) f(-\alpha(\mathbf{a}^k)+N_k)}}{e^{-\beta\alpha(\mathbf{a}^k) f(-\alpha(\bar{\mathbf{a}}^k)+N_k)}} \tag{40}$$

$$= \frac{1}{N} \sum_{k=1}^{N} \mathbf{T} \cdot \left(\frac{1}{\Gamma} \prod_{j=1}^{\mathcal{C}(\mathbf{a})} e^{-\beta f(j-1)} \right) \cdot \frac{e^{-\beta\alpha(\bar{\mathbf{a}}^k) f(-\alpha(\mathbf{a}^k)+N_k)}}{e^{-\beta\alpha(\mathbf{a}^k) f(-\alpha(\bar{\mathbf{a}}^k)+N_k)}} \tag{41}$$

$$= \frac{1}{N} \sum_{k=1}^{N} \mathbf{T}_{(\bar{\mathbf{a}}^k,\mathbf{a}^{-k}),(\mathbf{a}^k,\mathbf{a}^{-k})} \cdot \mathbf{u_a} \cdot \frac{e^{-\beta\alpha(\bar{\mathbf{a}}^k) f(-\alpha(\mathbf{a}^k)+N_k)}}{e^{-\beta\alpha(\mathbf{a}^k) f(-\alpha(\bar{\mathbf{a}}^k)+N_k)}} \tag{42}$$

The fraction inside the sum in Eq. 42 can be simplified as follows leading to further simplification of Eq. 42:

$$\sum_{\mathbf{a}_q \neq \mathbf{a}} \mathbf{T}_{\mathbf{a}_q,\mathbf{a}} \mathbf{u}_{\mathbf{a}_q} = \frac{1}{N} \sum_{k=1}^{N} \mathbf{T}_{(\bar{\mathbf{a}}^k,\mathbf{a}^{-k}),(\mathbf{a}^k,\mathbf{a}^{-k})} \cdot \mathbf{u_a} \cdot e^{sign(\mathbf{a}^k)\beta f(N_k)} \tag{43}$$

In Eq. 43 we can replace the element of the transition matrix $\mathbf{T}_{(\bar{\mathbf{a}}^k,\mathbf{a}^{-k}),(\mathbf{a}^k,\mathbf{a}^{-k})}$ by using the following:

$$\mathbf{T}_{(\bar{\mathbf{a}}^k,\mathbf{a}^{-k}),(\mathbf{a}^k,\mathbf{a}^{-k})} = \frac{1}{1 + e^{sign(\mathbf{a}^k)\beta f(N_k)}} \tag{44}$$

Using the expression for the transition matrix element from Eq. 44 into Eq. 43 and by using Eq. 34, we can simplify further:

$$\sum_{\mathbf{a}_q \neq \mathbf{a}} \mathbf{T}_{\mathbf{a}_q,\mathbf{a}} \mathbf{u}_{\mathbf{a}_q} = \frac{\mathbf{u_a}}{N} \sum_{k=1}^{N} \frac{1}{1 + e^{sign(\mathbf{a}^k)\beta f(N_k)}} \cdot e^{sign(\mathbf{a}^k)\beta f(N_k)} \tag{45}$$

$$= \frac{\mathbf{u_a}}{N} \sum_{k=1}^{N} \frac{1}{1 + e^{sign(\bar{\mathbf{a}}^k)\beta f(N_k)}} \tag{46}$$

$$= \mathbf{u_a} - \mathbf{u_a} \mathbf{T}_{\mathbf{a},\mathbf{a}} \tag{47}$$

The final step in the previous simplification shows that Eq. 21 holds for any $\mathbf{a} \in \{\mathrm{C}, \mathrm{D}\}^N$. Therefore, the candidate distribution we propose in Eq. **??** is the unique stationary distribution of the symmetric N-player game with two strategies. $\square$

Proof. **Proof of Corollary ??**

To show this result we count how many states are identical to a state $\mathbf{a} \in \{\mathrm{C}, \mathrm{D}\}^N$ in a symmetric game. When players are symmetric in a two-strategy game, states can be enumerated by counting the number of C players in that state. This can also be confirmed by the expression of the stationary distribution in Eq. **??**. Two distinct states $\mathbf{a}, \mathbf{a}'$ having the same number of cooperators (i.e., $\mathcal{C}(\mathbf{a}') = \mathcal{C}(\mathbf{a})$), have the same stationary distribution probability (i.e., $\mathbf{u}_{\mathbf{a}'} = \mathbf{u}_{\mathbf{a}}$).

In a game with N players, there can be k players playing C in exactly $\binom{N}{k}$ ways. As argued before, all of these states are identical and are also equiprobable in the stationary distribution. Therefore, the stationary distribution probability of having k, C players, $\mathbf{u}_k$, is,

$$\mathbf{u}_k = \sum_{\forall \mathbf{a}, \mathcal{C}(\mathbf{a})=k} \mathbf{u}_{\mathbf{a}} = \frac{1}{\Gamma}\binom{N}{k}\prod_{j=1}^{k} e^{-\beta f(j-1)} \tag{48}$$

Where the normalization factor Γ can also be simplified as:

$$\Gamma = \sum_{k=0}^{N}\binom{N}{k}\prod_{j=1}^{k} e^{-\beta f(j-1)} \tag{49}$$

□

References