Non-Vertical Cultural Transmission, Assortment, and the Evolution of Cooperation

immediate

July 18, 2021

This work was carried out under the supervision of **Dr. Yoav Ram** from the School of Zoology, Faculty of Life Sciences, Tel Aviv University.

Abstract

Cultural evolution of cooperation under vertical and non-vertical cultural transmission is studied, and conditions are found for fixation and coexistence of cooperation and defection. The evolution of cooperation is facilitated by its horizontal transmission and by an association between social interactions and horizontal transmission. The effect of oblique transmission depends on the horizontal transmission bias. Stable polymorphism of cooperation and defection can occur, and when it does, reduced association between social interactions and horizontal transmission evolves, which leads to a decreased frequency of cooperation and lower population mean fitness. The deterministic conditions are compared to outcomes of stochastic simulations of structured populations. Parallels are drawn with Hamilton's rule incorporating relatedness and assortment.

Contents

1 Introduction

Cooperative behavior can reduce an individual's fitness and increase the fitness of its conspecifics or competitors [1]. Nevertheless, cooperative behavior appears to occur in many animals [2], including humans, primates [3], rats [4], birds [5, 6], and lizards [7]. Evolution of cooperative behavior has been an important focus of research in evolutionary theory since at least the 1930s [8]. Since the work of Hamilton [9] and Axelrod and Hamilton [1], theories for the evolution of cooperative and altruistic behaviors have been intertwined often under the rubric of *kin selection*. Kin selection theory posits that natural selection is more likely to favor cooperation between more closely related individuals. The importance of *relatedness* to the evolution of cooperation and altruism was demonstrated by Hamilton [9], who showed that an allele that determines cooperative behavior will increase in frequency if the reproductive cost to the actor that cooperates, c, is less than the benfit to the recipient, b, times the relatedness, r, between the recipient and the actor. This condition is known as *Hamilton's rule*:

$$c < b \cdot r, \tag{1}$$

where the relatedness coefficient r measures the probability that an allele sampled from the cooperator is identical by descent to one at the same locus in the recipient.

There is an ongoing debate about to what extent kin selection explains evolution of cooperation and altruism. It has been suggested that kin selection to explain the cooperative behaviour of eusocial insects like the honey bee. The most significant argument against kin selection is that cooperation can evolve with zero relatedness [10]. This makes Hamilton's rule incomplete according to Wilson [10]. Foster et al. [11] reject this claim. They argue that altruism without relatedness can not evolve. They refer us to Hamilton who claimed that relatedness can arise without recent common ancestry. Wilson also criticises kin selection on the grounds that environmental or ecological factors probably be more important than relatedness in determining social actions. On the other hand, Foster et al. [11] argue that kin selection does not ignore ecology. Hamilton's rule shows that environmental factors causing a high benefit: cost ratio will favour cooperation.

Beside kin selection, two other major theroies were suggested to explain to evolution of cooperation.

Reciprocity suggests repeating interactions or individual recognition as key factors for explaining the evolution of cooperation. In *direct reciprocity* there are a repeated encounters between the same two individuals. In every encounter, each player has a choice between cooperation and defection. If I cooperate now, you may cooperate later. Hence, it may pay off to cooperate. This game-theoretic framework is known as the *repeated Prisoner's Dilemma*. Direct reciprocity can only lead to the evolution of cooperation if the cost is smaller than w the probability for another encounter between the same two individuals multiplied by the benefit,

$$c < b \cdot w\,. \tag{2}$$

Direct reciprocity assumes that both players are in a position to cooperate. Direct reciprocity can not explain cooperation in asymmetric interactions[12]. In humans, such interactions happen often, for example, humans donate money. *Indirect reciprocity* has been suggested to explain this behavior. Nowak [13] claims that direct reciprocity is like a barter economy based on the immediate exchange of goods, while indirect reciprocity resembles the invention of currency. The currency that "fuels the engines" of indirect reciprocity is *reputation*. However, reciprocity assumes repeating interactions and therefore, has difficulty explaining evolution of cooperation if no repeating interactions occurs.

Group Selection theory posits that cooperation is favoured because of the advantage to the whole group, if selection acts at the group level in addition to the individual level. A common model for group selection work as is: the population is divided into groups. In each group there are cooperators, which help other group members, and defectors, which do not help. Individuals reproduce proportional to their fitness. Offspring are added to the same group. If a group reaches a certain size it can split to two groups. A group that grows faster will split more often. Groups of cooperators grow faster than groups of defectors. Therefore, cooperation can evolve in this model when

$$\frac{b}{c} > 1 + \frac{n}{m}, \tag{3}$$

where n is the maximum group size and m is the number of groups.

All three theories mentioned above assume that cooperation is genetically determined. This raise the question, is it possible that cooperation is determined by environmental or social influences. Cooperative behavior may be subject to *cultural transmission*, which allows an individual to acquire attitudes or behavioral traits from other individuals in its social group through imitation, learning, or other modes of communication [14, 15]. Cultural transmission may be modeled as vertical, horizontal, or oblique: vertical transmission occurs between parents and offspring, horizontal transmission occurs between individuals from the same generation, and oblique transmission occurs to offspring from the generation to which their parents belong (i.e. from non-parental adults). Evolution under either of these transmission models can be be more rapid than under pure vertical transmission [14, 16, 17].

Eshel and Cavalli-Sforza [18] studied a related model for the evolution of cooperative behavior. Their model included *assortative meeting*, or non-random encounters, where a fraction m of individuals in the population each interact specifically with an individual of the same phenotype, and a fraction $1-m$ interacts with a randomly chosen individual. Such assortative meeting may be due, for example, to population structure or active partner choice. In their model, cooperative behavior can evolve if [18, eq. 3.2]

$$c < b \cdot m, \tag{4}$$

where b and c are the benefit and cost of cooperation[1].

The role of assortment in the evolution of altruism was emphasized by Fletcher and Doebeli [19]. They found that in a *public-goods* game, altruism will evolve if cooperative individuals experience more cooperation, on average, than defecting individuals, and “thus, the evolution of altruism requires

[1] In an extended model, which allows an individual to encounter N individuals before choosing a partner, the right hand side is multiplied by $E[N]$, the expected number of encounters [18, eq. 4.6].

(positive) assortment between focal *cooperative* players and cooperative acts in their interaction environment." With some change in parameters, this condition is summarized by [19, eq. 2.3]

$$c < b \cdot (p_C - p_D) , \tag{5}$$

where p_C is the probability that a cooperator receives help, and p_D is the probability that a defector receives help[2]. Bijma and Aanen [20] obtained a result related to inequality 5 for other games.

Cooperation can also evolve when interactions are determined by population structure. For example, Ohtsuki et al. [21] studied populations on graphs with average degree k, that is, the average individual has k potential interaction partners. Assuming that selection is weak and that the population size is much larger than k (i.e. sparse structure), they found that cooperative behaviour can evolve if [21]

$$c < b \cdot \frac{1}{k} . \tag{6}$$

They thus interpret $1/k$ as *social relatedness* or *social viscosity* [21].

Feldman et al. [22] introduced the first model for the evolution of altruism by cultural transmission with kin selection and demonstrated that if the fidelity of cultural transmission of altruism is φ, then the condition for evolution of altruism in the case of sib-to-sib altruism is [22, Eq. 16]

$$c < b \cdot \varphi - \frac{1 - \varphi}{\varphi} . \tag{7}$$

In inequality 7, φ replaces relatedness (r in inequality 1) or assortment (m in inequality 4), but the effective benefit $b \cdot \varphi$ is reduced by $(1-\varphi)/\varphi$. This shows that under a cultural transmission, the condition for the evolutionary success of altruism entails a modification of Hamilton's rule (inequality 1).

Both Woodcock [23] and Lewin-Epstein et al. [24] demonstrated that non-vertical transmission can help explain the evolution of cooperative behavior, the former using simulations with cultural transmission, the latter using a model where cooperation is mediated by host-associated microbes. Indeed, models in which microbes affect their host's behavior [24, 25, 26] are mathematically similar to models of cultural transmission, and they also emphasize the role of non-vertical transmission [14].

Handley and Mathew [27] studied the importance of culture on human behavior. They showed that the probablity of indvidual to cooperate with unrelated strangers from a different group in transient interactions corresponds to the degree of cultural similarity between those groups. Therefore, they have suggested that group-level selection on culturally differentiated populations can explain cooperation between unrelated humans from different groups.

To understand the evolution of cooperation we are going to use *replicator dynamics*. The replicator in replicator dynamics has the ability to make one or more copies of itself. The replicator can be a gene, a phenotype, a strategy in a game and etc. In evolutionary game theory context replicator is a different strategy in the game. For example, in the case of cooperation whether the individual

[2]Inequality 5 generalizes inequalities 1 and 4 by substituting $p_C = r+p$, $p_D = p$ and $p_C = m+(1-m)p$, $p_D = (1-m)p$, respectively, where p is the frequency of cooperators.

is a cooperative or a defector. In replicator dynamics we assume large population of replicators, which interact with respect to their frequency. Interactions of different replicator affect the fitness acorrding to some payoff matrix. This payoff matrix depands on the game which is played. The most common game to describe cooperation is the *Prisoner's Dilemma*[28]. Similar to dominant strategies bringing forth Nash equilibria when games are repeated, strategies in replicator dynamics can become evolutionary stable. Such strategies are called *evolutionarily stable strategies (ESS)*. Such strategies cannot be invaded by any other strategy that is initially rare. Evolutionarily stable strategies maximize the expected fitness of its replicators and therefore, maximize the population mean fitness. A fundamental question is to test if natural selection, operating within the framework of known genetical models, leads to ESS. The answer to the question above is that natural selection in many cases does not lead to an increase in the population mean fitness. Therefore, it is not generally true that natural selection does operate as to maximize the average individual fitness resulting in an ESS. Sacks [29] even showed that natural selection at a diploid locus model, when relative fitnesses of the different genotypes are functions of gene frequency, leads to stable equilibrium gene frequencies that corresponds to not maximum mean fitness. Instead of searching for evolutionarily stable strategies, we focus on *evolutionary genetic stblility (EGS)*[30]. EGS occurs when the frequencies of types (genotypes, phenotypes etc.) remains stable. Here, we study both *local stability* and *external stability*. *Local stability* occurs when a system near the equilibrium wiil approach it. *External stability* [31, 32] is local stability of the equilibrium to invasion by a modifier allele m. One of the main questions in the evolution of cooperation is under what conditions such invasions are possible, or in other words under what conditions the system is externally unstable?

Here, we study models for the cultural evolution of cooperation that include both vertical and non-vertical transmission. In our models behavioral changes are mediated by cultural transmission that can occur specifically during social interactions. For instance, there may be an association between the choice of partner for social interaction and the choice of partner for cultural transmission, or when an individual interacts with an individual of a different phenotype, exposure to the latter may lead the former to convert its phenotype. Our results demonstrate that cultural transmission, when associated with social interactions, can enhance the evolution of cooperation even when genetic transmission cannot, partly because it facilitates the generation of assortment [19], and partly because it diminishes the effect of selection (due to non-vertical transmission from non-reproducing individuals [17]).

2 Models

2.1 Basic Model

Consider a large population whose members can have one of two phenotypes: $\phi = A$ for cooperators or $\phi = B$ for defectors. An offspring inherits its phenotype from its parent via vertical transmission with probability v or from a random individual in the parental population via oblique transmission with probability $(1-v)$ (Figure 1a). Following Ram et al. [17], given that the parent's phenotype is ϕ and assuming uni-parental inheritance [33], the conditional probability that the phenotype ϕ' of the offspring is A is

$$P(\phi' = A \mid \phi) = \begin{cases} v + (1-v)p, & \text{if } \phi = A \\ (1-v)p, & \text{if } \phi = B \end{cases}, \tag{8}$$

where $p = P(\phi = A)$ is the frequency of A among all adults in the parental generation.

Not all adults become parents, and we denote the frequency of phenotype A among parents by $\dot{p}$. Therefore, the frequency $\hat{p}$ of phenotype A among juveniles (after selection and vertical and oblique transmission) is

$$\hat{p} \ = \dot{p}[v + (1-v)p] + (1-\dot{p})[(1-v)p] \ = v\dot{p} + (1-v)p \ . \tag{9}$$

Individuals are assumed to interact according to a *Prisoner's Dilemma*. Specifically, individuals interact in pairs; a cooperator suffers a fitness cost $0 < c < 1$, and its partner gains a fitness benefit b, where we assume $c < b$. Figure 1a shows the payoff matrix, i.e. the fitness of an individual with phenotype ϕ_1 when interacting with a partner of phenotype ϕ_2. The choise of Prisoner's Dilemma as the interaction model was motivated by the fact that prisoner's dilemma is a common game used to study evolution of cooperation[1], [34], [28]. Although we decided to focus on Prisoner's Dilemma, other games such as stag hunt[35] may be a better explaination of cooperation behavior in humman[36].

Social interactions occur randomly: two juvenile individuals with phenotype A interact with probability $\hat{p}^2$, two juveniles with phenotype B interact with probability $(1-\hat{p})^2$, and two juveniles with different phenotypes interact with probability $2\hat{p}(1-\hat{p})$. Horizontal cultural transmission occurs between pairs of individuals from the same generation. It occurs between socially interacting partners with probability α, or between a random pair with probability $1-\alpha$ (see Figure 1b). However, horizontal transmission is not always successful, as one partner may reject the other's phenotype. The probability of successful horizontal transmission of phenotypes A and B are T_A and T_B, respectively (Table 1, Figure 1d). Thus, the frequency p' of phenotype A among adults in the next generation, after horizontal transmission, is

$$\begin{aligned} p' \ &= \hat{p}^2\big[\alpha + (1-\alpha)\big(\hat{p} + (1-\hat{p})(1-T_B)\big)\big] \ + \\ &\quad \hat{p}(1-\hat{p})\big[\alpha(1-T_B) + (1-\alpha)\big(\hat{p} + (1-\hat{p})(1-T_B)\big)\big] \ + \\ &\quad (1-\hat{p})\hat{p}\big[\alpha T_A + (1-\alpha)\hat{p}T_A\big] \ + (1-\hat{p})^2\big[(1-\alpha)\hat{p}T_A\big] \\ &= \hat{p}^2(T_B - T_A) \ + \ \hat{p}(1 + T_A - T_B) \ . \end{aligned} \tag{10}$$

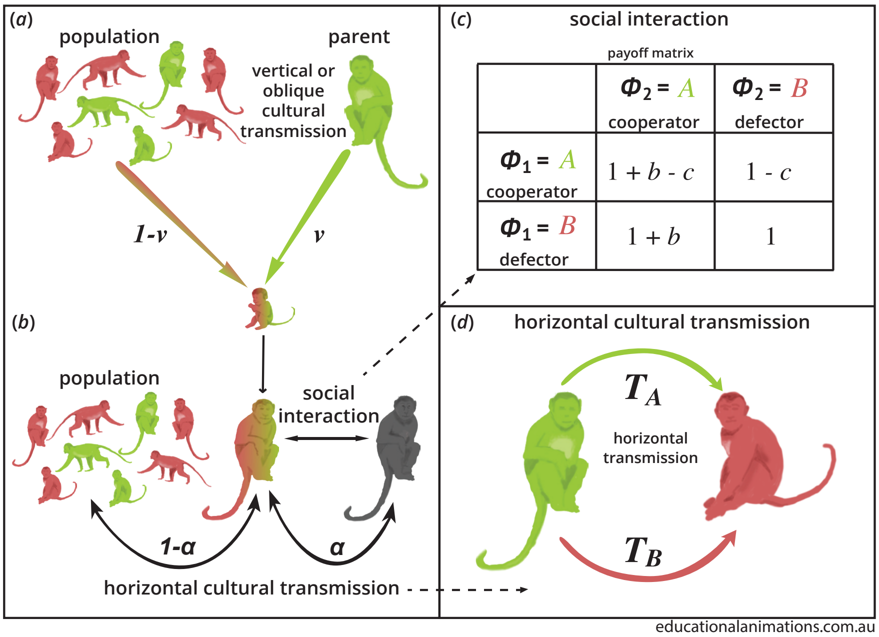

Figure 1: Model illustration. **(a)** First, offspring inherit their parent's phenotype via vertical cultural transmission with probability v, or the phenotype of a random non-parental adult via oblique cultural transmission with probability $1 - v$. **(b)** Second, adults socially interact in pairs in a Prisoner's Dilemma game. Horizontal cultural transmission occurs from a random individual in the population, with probability $1 - \alpha$, or from the social partner, with probability α, where α is the interaction-transmission association parameter. **(c)** The Prisoner's Dilemma payoff matrix shows the fitness of phenotype ϕ_1 when interacting with phenotype ϕ_2. **(d)** The probabilities of successful horizontal cultural transmission of phenotypes A (cooperator) and B (defector) are T_A and T_B, respectively.

The frequency of A among parents (i.e. after selection) follows a similar dynamic, but also includes the effect of natural selection, and is therefore

$$\begin{aligned}\bar{w}\dot{p}' =&\hat{p}^2(1+b-c)\big[\alpha+(1-\alpha)\big(\hat{p}+(1-\hat{p})(1-T_B)\big)\big] + \\ &\hat{p}(1-\hat{p})(1-c)\big[\alpha(1-T_B)+(1-\alpha)\big(\hat{p}+(1-\hat{p})(1-T_B)\big)\big] + \\ &(1-\hat{p})\hat{p}(1+b)\big[\alpha T_A+(1-\alpha)\hat{p}T_A\big] \ + \ (1-\hat{p})^2\big[(1-\alpha)\hat{p}T_A\big] \ ,\end{aligned} \tag{11}$$

where fitness values are taken from Figure 1c and Table 1, and the population mean fitness is $\bar{w} = 1+\hat{p}(b-c)$. Starting from Eq. 9 with $\hat{p}' = v\dot{p}' + (1-v)p'$, we substitute p' from Eq. 10 and $\dot{p}'$ from Eq. 11 and obtain

$$\begin{aligned}\hat{p}' =&\frac{v}{\bar{w}}\left[\hat{p}^2(1+b-c)\Big(1-(1-\hat{p})(1-\alpha)T_B)\Big)\right] + \\ &\frac{v}{\bar{w}}\left[\hat{p}(1-\hat{p})(1-c)\big(\hat{p}(1-\alpha)T_B+1-T_B\big)\right] + \\ &\frac{v}{\bar{w}}\left[\hat{p}(1-\hat{p})(1+b)\big(\hat{p}(1-\alpha)+\alpha\big)T_A\right] + \\ &\frac{v}{\bar{w}}(1-\hat{p})^2\hat{p}(1-\alpha)T_A \ + \ (1-v)\hat{p}^2(T_B-T_A) \ + \ (1-v)\hat{p}(1+T_A-T_B) \ .\end{aligned} \tag{12}$$

Table 2 lists the model variables and parameters.

2.2 Invasion Model

In this section, we extend the model described in subsection 2.1. We assume coexistence of both phenotypes (cooperators and defectors) and that the frequency of each phenotye stays stable over time. Hence, we assume that the we start from a stable equilibrium. Denote the frequency of cooperators in the population as $\hat{p}^*$ and frequency of defectors as $1-\hat{p}^*$.

We investigate the evolution of interaction-transmission association α. We assume that the initial population has interaction-transmission association α_1. What would happen to cooperators frequency if a new mutation would occur that changed the interaction-transmission association to a different value α_2 without direct effect on fitness? Here, the invaders are the individuals with interaction-transmission association α_2. Figure 2 illustrates the invasion.

Table 1: Interaction frequency, fitness, and transmission probabilities.

Phenotype ϕ_1	Phenotype ϕ_2	Frequency	Fitness of ϕ_1	$P(\phi_1 = A)$ via horizontal transmission:	
				from partner, α	from population, $(1-\alpha)$
A	A	$\hat{p}^2$	$1+b-c$	1	$\hat{p} + (1-\hat{p})(1-T_B)$
A	B	$\hat{p}(1-\hat{p})$	$1-c$	$1-T_B$	$\hat{p} + (1-\hat{p})(1-T_B)$
B	A	$\hat{p}(1-\hat{p})$	$1+b$	T_A	$\hat{p}T_A$
B	B	$(1-\hat{p})^2$	1	0	$\hat{p}T_A$

Table 2: Model variables and parameters.

Symbol	Description	Values
A	Cooperator phenotype	
B	Defector phenotype	
p	Frequency of phenotype A among adults	$[0, 1]$
$\dot{p}$	Frequency of phenotype A among parents	$[0, 1]$
$\hat{p}$	Frequency of phenotype A among juveniles	$[0, 1]$
v	Vertical transmission rate	$[0, 1]$
c	Cost of cooperation	$(0, 1)$
b	Benefit of cooperation	$c < b$
α	Probability of interaction-transmission association	$[0, 1]$
T_A, T_B	Horizontal transmission rates of phenotype A and B	$(0, 1)$

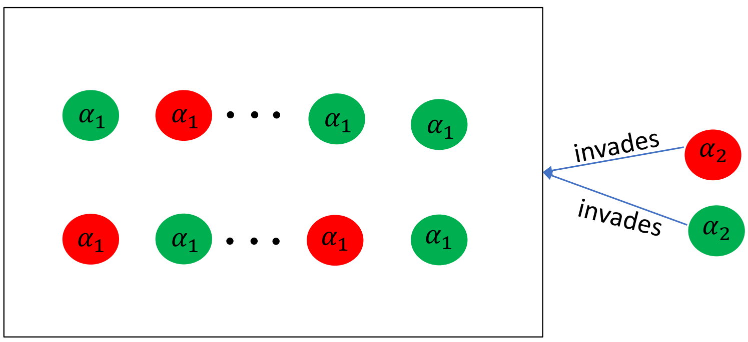

Figure 2: Invasion model illustration. Starting from stable equilibrium with coexistence of both phenotypes. A mutation that affected only the interaction-transmission association α invades the population

2.3 Population structure model

Interaction-transmission association may also emerge from population structure. Consider a population colonizing a two-dimensional grid of size 100-by-100, where each site is inhabited by one individual, similarly to the model of Lewin-Epstein and Hadany [25]. Each individual is characterized by its phenotype: either cooperator, A, or defector, B. Initially, each site in the grid is randomly colonized by either a cooperator or a defector, with equal probability. In each generation, half of the individuals are randomly chosen to "initiate" interactions, and these initiators interact with a random neighbor (i.e. individual in a neighboring site) in a Prisoner's Dilemma game (Figure 1c) and a random neighbor (with replacement) for horizontal cultural transmission (Figure 1b). The expected number of each of these interactions per individual per generation is one, but the realized number of interactions can be zero, one, or even more than one, and in every interaction both individuals are affected, not just the initiator. The effective interaction-transmission association α in this model is the probability that the same neighbor is picked for both interactions, or $\alpha = 1/M$, where M is the number of neighbors. On an infinite grid, $M = 8$ (i.e. Moore neighbourhood [37]), but on a finite grid M can be lower in neighbourhoods close to the grid border. As before, T_A and T_B are the probabilities of successful horizontal transmission of phenotypes A and B, respectively.

The order of the interactions across the grid at each generation is random. After all interactions take place, an individual's fitness is determined by $w = 1+b\cdot n_b-c\cdot n_c$, where n_b is the number of interactions that individual had with cooperative neighbors, and n_c is the number of interactions in which that individual cooperated (note that the phenotype may change between consecutive interactions due to horizontal transmission). Then, a new generation is produced, and the sites can be settled by offspring of any parent, not just the neighboring parents. Selection is global, rather than local, in accordance with our deterministic model: The parent is randomly drawn with probability proportional to its fitness, divided by the sum of the fitness values of all potential parents. Offspring are assumed to have the same phenotype as their parents (i.e. $v = 1$).

3 Results

We determine the equilibria of the model in Eq. 12 and analyze their local stability. We then analyze the evolution of a modifier of interaction-transmission association, α. Finally, we compare derived conditions to outcomes of stochastic simulations with a structured population.

3.1 Evolution of cooperation

To learn about the evolution of cooperation we investigate local stability of the equilibria of the model in Eq. 12. The equilibria are the solutions of $\hat{p}' - \hat{p} = 0$. Note that Eq. 12 may look simple cubic polynomial. However, because $\bar{w}$ is a function of $\hat{p}$, Eq. 12 is not simple polynomial but a fractional polynomial. The solution of fractional polynomial is not trivial and that is why it is better to find the equilibria and analyze the stability. Let $f(\hat{p}) = \bar{w}(\hat{p}' - \hat{p})$. Then, using *SymPy* [38], a Python library for symbolic mathematics, this simplifies to

$$f(\hat{p}) = \bar{w}(\hat{p}' - \hat{p}) = \beta_1 \hat{p}^3 + \beta_2 \hat{p}^2 + \beta_3 \hat{p} \ , \tag{13}$$

where

$$\begin{aligned} \beta_1 &= \big[c(1-v) - b(1-\alpha v)\big](T_A - T_B) \ , \\ \beta_2 &= -\beta_1 - \beta_3 \ , \\ \beta_3 &= \alpha b v T_A - cv(1-T_B) + (T_A - T_B) \ . \end{aligned} \tag{14}$$

If $T = T_A = T_B$ then $\beta_1 = 0$ and $\beta_3 = -\beta_2 = \alpha b v T - cv(1-T)$, and $f(\hat{p})$ becomes a quadratic polynomial,

$$f(\hat{p}) = \hat{p}(1-\hat{p})\big[\alpha b v T - cv(1-T)\big] \ . \tag{15}$$

Clearly the only two equilibria are the fixations $\hat{p} = 0$ and $\hat{p} = 1$, which are are locally stable if $f'(\hat{p}) < 0$ near the equilibrium (see Appendix A), where $f'(\hat{p}) = (1-2\hat{p})\big[\alpha b v T - cv(1-T)\big]$, so that

$$\begin{aligned} f'(0) &= \alpha b v T - cv(1-T) \ , \\ f'(1) &= -\alpha b v T + cv(1-T) \ . \end{aligned} \tag{16}$$

In the general case where $T_A \neq T_B$, the coefficient β_1 is not necessarily zero, and $f(\hat{p})$ is a cubic polynomial. Therefore, three equilibria may exist, two of which are $\hat{p} = 0$ and $\hat{p} = 1$, and the third is

$$\hat{p}^* = \frac{\beta_3}{\beta_1} = \frac{\alpha b v T_A - cv(1-T_B) + (T_A - T_B)}{\big[c(1-v) - b(1-\alpha v)\big](T_A - T_B)} \ . \tag{17}$$

Note that the sign of the cubic (Eq. 13) at positive (negative) infinity is equal (opposite) to the sign of β_1. If $T_A > T_B$, then

$$\beta_1 < [c(1-\alpha v) - b(1-\alpha v)](T_A - T_B) = (1-\alpha v)(c-b)(T_A - T_B) < 0 \ , \tag{18}$$

since $c < b$ and $\alpha v < 1$. Hence the signs of the cubic at positive and negative infinity are negative and positive, respectively. First, if $\beta_3 < \beta_1$ then $1 < \hat{p}^*$. Also, $f'(0) < 0$ and $f'(1) > 0$; that is, fixation of the defector phenotype B is the only locally stable feasible equilibrium. Second, if $\beta_1 < \beta_3 < 0$ then $0 < \hat{p}^* < 1$ and therefore $f'(0) < 0$ and $f'(1) < 0$ so that both fixations are locally stable and $\hat{p}^*$ separates the domains of attraction. Third, if $0 < \beta_3$ then $\hat{p}^* < 0$ and therefore $f'(0) > 0$ and $f'(1) < 0$; that is, fixation of the cooperator phenotype A is the only locally stable legitimate equilibrium.

Similarly, if $T_A < T_B$, then

$$\beta_1 > [c(1-\alpha v) - b(1-\alpha v)](T_A - T_B) = (1-\alpha v)(c-b)(T_A - T_B) > 0\,, \tag{19}$$

since $c < b$ and $\alpha v < 1$, and the signs of the cubic at positive and negative infinity are positive and negative, respectively. First, if $\beta_3 < 0$ then $\hat{p}^* < 0$ and therefore $f'(0) < 0$ and $f'(1) > 0$; that is, fixation of the defector phenotype A is the only locally stable legitimate equilibrium. Second, if $0 < \beta_3 < \beta_1$ then $0 < \hat{p}^* < 1$ and therefore $f'(0) > 0$ and $f'(1) > 0$; that is, both fixations are locally unstable and $\hat{p}^*$ is a stable polymorphic equilibrium. Third, if $\beta_1 < \beta_3$ then $\hat{p}^* > 1$ and therefore $f'(0) > 0$ and $f'(1) < 0$, and fixation of the cooperator phenotype A is the only locally stable feasible equilibrium.

This analysis can be summarized as follows:

1. *Fixation of cooperation*: if *(i)* $T = T_A = T_B$ and $c < b \cdot \frac{\alpha T}{1-T}$; or if *(ii)* $T_A > T_B$ and $0 < \beta_3$; or if *(iii)* $T_A < T_B$ and $\beta_1 < \beta_3$.

2. *Fixation of the defection*: if *(iv)* $T = T_A = T_B$ and $c > b \cdot \frac{\alpha T}{1-T}$; or if *(v)* $T_A > T_B$ and $\beta_3 < \beta_1 < 0$; or if *(vi)* $T_A < T_B$ and $\beta_3 < 0$.

3. *polymorphism of both phenotypes at* $\hat{p}^*$: if *(vii)* $T_A < T_B$ and $0 < \beta_3 < \beta_1$.

4. *Fixation of either phenotype depending on initial frequency*: if *(viii)* $T_A > T_B$ and $\beta_1 < \beta_3 < 0$.

Define the following cost thresholds, γ_1 and γ_2, and the vertical transmission threshold, $\hat{v}$,

$$\gamma_1 = \frac{bv\alpha T_A + (T_A - T_B)}{v(1-T_B)}, \quad \gamma_2 = \frac{bv\alpha T_B + (1+b)(T_A - T_B)}{v(1-T_B) + (1-v)(T_A - T_B)}, \quad \hat{v} = \frac{T_B - T_A}{1 - T_A}\,. \tag{20}$$

We now proceed to use the cost thresholds, γ_1 and γ_2, and the vertical transmission threshold, $\hat{v}$ (Eq. 20). First, assume $T_A < T_B$. $\beta_3 < 0$ requires $\gamma_1 < c$. For $\beta_3 < \beta_1$ we need $c\big[v(1-T_B)+(1-v)(T_A-T_B)\big] > bv\alpha T_B + (1+b)(T_A - T_B)$. Note that the expression in the square brackets is positive if and only if $v > \hat{v}$. Thus, for $\beta_3 < \beta_1$ we need $v > \hat{v}$ and $\gamma_2 < c$ or $v < \hat{v}$ and $c < \gamma_2$, and for $0 < \beta_3 < \beta_1$ we need $v > \hat{v}$ and $\gamma_2 < c < \gamma_1$, or $v < \hat{v}$ and $c < \min(\gamma_1, \gamma_2)$. For $\beta_1 < \beta_3$ we need $v > \hat{v}$ and $c < \gamma_2$ or $v < \hat{v}$ and $\gamma_2 < c$. However, some of these conditions cannot be met, since $v < \hat{v}$ implies $c < 1 < \gamma_2$.

Second, assume $T_A > T_B$. $\beta_3 > 0$ requires $\gamma_1 > c$. For $\beta_1 < \beta_3$ we need $c\big[v(1-T_B)+(1-v)(T_A-T_B)\big] < bv\alpha T_B + (1+b)(T_A - T_B)$. Thus for $\beta_1 < \beta_3$ we need $v > \hat{v}$ and $c < \gamma_2$ or $v < \hat{v}$ and $c > \gamma_2$. But $\hat{v} < 0$ when $T_A > T_B$, and therefore we have $\beta_1 < \beta_3$ if $c < \gamma_2$. Similarly, we have $\beta_3 < \beta_1$ if $c > \hat{\gamma}_2$.

Then we have the following result.

Result 1. *With vertical, horizontal, and oblique transmission, the cultural evolution of a cooperation follows one of the following scenarios in terms of the cost thresholds γ_1 and γ_2 and the vertical transmission threshold $\hat{v}$ (Eq. 20):*

1. Fixation of cooperation*: if* (i) $T_A \geq T_B$ *and* $c < \gamma_1$*; or if* (ii) $T_A < T_B$ *and* $v > \hat{v}$ *and* $c < \gamma_2$.
2. Fixation of defection*: if* (iii) $T_A \geq T_B$ *and* $\gamma_2 < c$*; or if* (iv) $T_A < T_B$ *and* $\gamma_1 < c$.
3. Stable polymorphism*: if* (v) $T_A < T_B$ *and* $v < \hat{v}$ *and* $c < \gamma_1$*; or if* (vi) $T_A < T_B$ *and* $v > \hat{v}$ *and* $\gamma_2 < c < \gamma_1$.
4. Unstable polymorphism*: if* (vii) $T_A > T_B$ *and* $\gamma_1 < c < \gamma_2$.

Thus, cooperation can take over the population if it has either a horizontal transmission advantage, or if it has a horizontal transmission disadvantage, but the vertical transmission rate is high enough. In either case, the cost of cooperation must be small enough. A stable polymorphism can exist between cooperation and defection only if defection has a horizontal transmission advantage. In this case, the existence of a stable polymorphism depends on the interplay between the benefit and cost of cooperation and the vertical transmission rate. These conditions are illustrated in Figures 3a, 3b, 4a, and 4b. Note that *stable* and unstable polymorphism are also called, respectively, *coexistence* and *bistable competition*.

Much of the literature on evolution of cooperation focuses on conditions for an initially rare cooperative phenotype to invade a population of defectors. The following remarks address this condition.

Remark 1. *If the initial frequency of cooperation is very close to zero, then its frequency will increase if the cost of cooperation is low enough,*

$$c < \gamma_1 = \frac{bv\alpha T_A + (T_A - T_B)}{v(1 - T_B)} \ . \tag{21}$$

This unites the conditions for fixation of cooperation and for stable polymorphism, both of which entail instability of the state where defection is fixed, $\hat{p} = 0$.

Importantly, increasing interaction-transmission association α increases the cost threshold ($\partial\gamma_1/\partial\alpha > 0$), making it easier for cooperation to increase in frequency when initially rare. Similarly, increasing the horizontal transmission of cooperation, T_A, increases the threshold ($\partial\gamma_1/\partial T_A > 0$), facilitating the evolution of cooperation ((Figure 4a and 4b). However, increasing the horizontal transmission of defection, T_B, can increase or decrease the cost threshold, but it increases the cost threshold when the threshold is already above one ($c < 1 < \gamma_1$): $\partial\gamma_1/\partial T_B$ is positive when $T_A > \frac{1}{1+\alpha bv}$, which gives $\gamma_1 > 1/v$. Therefore, increasing T_B decreases the cost threshold and limits the evolution of cooperation, but only if $T_A < \frac{1}{1+\alpha bv}$.

Increasing the vertical transmission rate, v, can either increase or decrease the cost threshold, depending on the horizontal transmission bias, $T_A - T_B$, because $\text{sign}(\partial\gamma_1/\partial v) = -\text{sign}(T_A - T_B)$. When $T_A < T_B$ we have $\partial\gamma_1/\partial v > 0$, and as the vertical transmission rate increases, the cost threshold increases, making it easier for cooperation to increase when rare (Figure 3b). In contrast, when $T_A > T_B$ we get $\partial\gamma_1/\partial v < 0$, and therefore as the vertical transmission rate increases, the cost threshold decreases, making it harder for cooperation to increase when rare (Figure 3a).

In general, this condition cannot be formulated in the form of Hamilton's rule due to the bias in horizontal transmission, represented by $T_A - T_B$. If $T_A = T_B$, then, from Result 1 and inequality 21, cooperation will take over the population from any initial frequency if the cost is low enough,

$$c < b \cdot \frac{\alpha T}{1 - T} \,, \tag{22}$$

and regardless of the vertical transmission rate, v. This condition can be interpreted as a version of Hamilton's rule ($c < b \cdot r$, inequality 1) or as a version of inequality 5, where $\alpha T/(1 - T)$ can be regarded as the *effective relatedness* or *effective assortment*, respectively. Note that the right-hand side of inequality 22 equals γ_1 when $T = T_A = T_B$.

From inequality 21, without interaction-transmission association ($\alpha = 0$), cooperation will increase when it is rare if there is horizontal transmission bias for cooperation, $T_A > T_B$, and

$$c < \frac{T_A - T_B}{v(1 - T_B)} \,. \tag{23}$$

Figure 4a illustrates this condition (for $v = 1$), which is obtained by setting $\alpha = 0$ in inequality 21. In this case, the benefit of cooperation, b, does not affect the evolution of cooperation, and the outcome is determined only by cultural transmission. Further, inequality 21 shows that with perfect interaction-transmission association ($\alpha = 1$), cooperation will increase when rare if

$$c < \frac{bvT_A + (T_A - T_B)}{v(1 - T_B)} \,. \tag{24}$$

In the absence of oblique transmission, $v = 1$, the only equilibria are the fixation states, $\dot{p} = 0$ and $\dot{p} = 1$, and cooperation will evolve from any initial frequency (i.e. $\dot{p}' > \dot{p}$) if inequality 24 applies (Figure 4). This is similar to case of microbe-induced cooperation studied by Lewin-Epstein et al. [24]; therefore when $v = 1$, this remark is equivalent to their eq. 1.

It is interesting to examine the general effect of interaction-transmission association α on the evolution of cooperation. Define the interaction-transmission association thresholds, a_1 and a_2, as

$$a_1 = \frac{c \cdot v(1 - T_A) - (T_A - T_B)(1 + b - c)}{b \cdot v \cdot T_B}, \quad a_2 = \frac{c \cdot v(1 - T_B) - (T_A - T_B)}{b \cdot v \cdot T_A} \,. \tag{25}$$

Remark 2. *Cooperation will increase when rare if interaction-transmission association is high enough, specifically if $a_2 < \alpha$.*

Figures 3c and 3d illustrate this condition. With horizontal transmission bias for cooperation, $T_A > T_B$, cooperation can fix from any initial frequency if $a_2 < \alpha$ (green area in the figures). With horizontal bias favoring defection, $T_A < T_B$, cooperation can fix from any frequency if α is large enough, $a_1 < \alpha$ (green area with $T_A < T_B$), and can reach stable polymorphism if α is intermediate, $a_2 < \alpha < a_1$

(yellow area). Without horizontal bias, $T_A = T_B$, fixation of cooperation occurs if α is high enough, $\frac{c}{b} \cdot \frac{1-T}{T} < \alpha$ (inequality 22; in this case $a_1 = a_2$).

Interestingly, because $\text{sign}(\partial a_2/\partial v) = \text{sign}(T_A - T_B)$, the effect of the vertical transmission rate v on a_1 and a_2 depends on the horizontal transmission bias. That is, if $T_A > T_B$, then evolution of cooperation is facilitated by oblique transmission, whereas if $T_A < T_B$, then evolution of cooperation is facilitated by vertical transmission (Figures 3c and 3d).

Next, we examine the roles of vertical and oblique transmission in the evolution of cooperation. Fixation of cooperation is possible only if the vertical transmission rate is high enough,

$$v > \hat{v} = \frac{T_B - T_A}{1 - T_A} \ . \tag{26}$$

This condition is necessary for fixation of cooperation, but it is not sufficient. If horizontal transmission is biased for cooperation, $T_A > T_B$, cooperation can fix with any vertical transmission rate (because $\hat{v} < 0$). In contrast, if horizontal transmission is biased for defection, $T_A < T_B$, cooperation can fix only if the vertical transmission rate is high enough: in this case oblique transmission can prevent fixation of cooperation (see Figures 3b and 3d).

With only vertical transmission ($v = 1$), from inequality 21, cooperation increases when rare if

$$c < \frac{b\alpha T_A + (T_A - T_B)}{1 - T_B} \ , \tag{27}$$

which can also be written as

$$\frac{c(1 - T_B) - (T_A - T_B)}{bT_A} < \alpha \ . \tag{28}$$

In the absence of vertical transmission ($v = 0$), from recursion 12 we see that the frequency of the cooperator phenotype among adults increases every generation, i.e. $p' > p$, if there is a horizontal transmission bias in favor of cooperation, namely $T_A > T_B$. That is, if $v = 0$, then selection plays no role in the evolution of cooperation (i.e. b and c do not affect $\hat{p}'$). The dynamics are determined solely by differential horizontal transmission of the two phenotypes. With no bias in horizontal transmission, $T_A = T_B$, phenotype frequencies do not change, $\hat{p}' = \hat{p}$.

Cooperation and defection can coexist at frequencies $\hat{p}^*$ and $1 - \hat{p}^*$ (Eq. 17). When it is feasible, this equilibrium is stable or unstable under the conditions of Result 1, parts 3 and 4, respectively. The yellow and blue areas in Figures 4 and 3 show cases of stable and unstable polymorphism, respectively. When $\hat{p}^*$ is unstable, cooperation will fix if its initial frequency is $\hat{p} > \hat{p}^*$, and defection will fix if $\hat{p} < \hat{p}^*$ as shown in Figure 5a. $\hat{p}^*$ is unstable when there is horizontal transmission bias for cooperation, $T_A > T_B$, and the cost is intermediate, $\gamma_1 < c < \gamma_2$. Figure 4d shows $\hat{p}' - \hat{p}$ as a function of $\hat{p}$ and Figure 5 shows the frequency of cooperation over time in both stable and unstable equilibrium regimes.

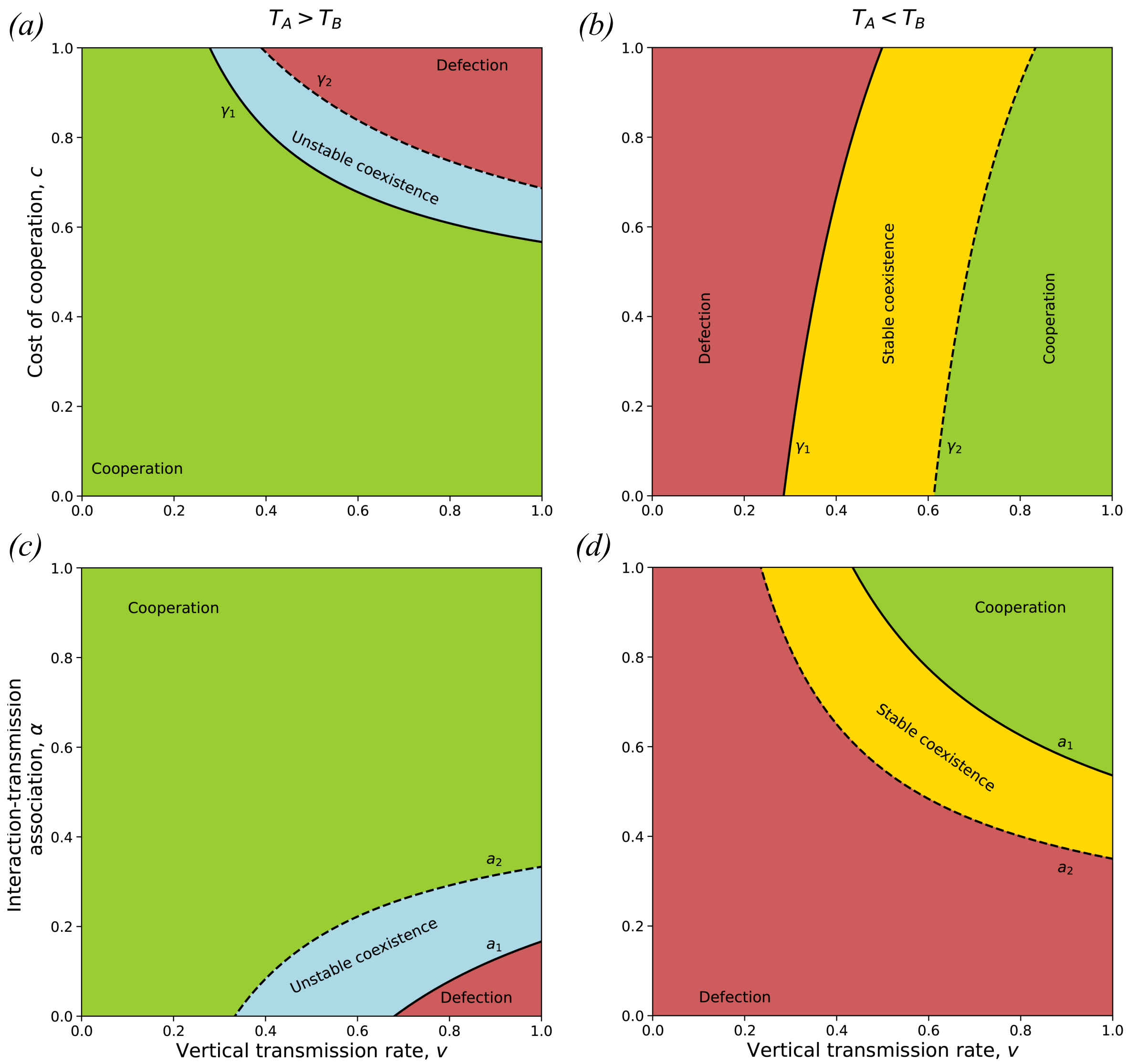

Figure 3: Evolution of cooperation under vertical, oblique, and horizontal cultural transmission. The figure shows parameter ranges for global fixation of cooperation (green), global fixation of defection (red), fixation of either cooperation or defection depending on the initial conditions, i.e. unstable polymorphism (blue), and stable polymorphism of cooperation and defection (yellow). In all cases the vertical transmission rate v is on the x-axis. (**a-b**) Cost of cooperation c is on the y-axis and the cost thresholds γ_1 and γ_2 (Eqs. 20) are represented by the solid and dashed lines, respectively. (**c-d**) Interaction-transmission association α is on the y-axis and the interaction-transmission association thresholds a_1 and a_2 (Eqs. 25) are represented by the solid and dashed lines, respectively. Horizontal transmission is biased in favor of cooperation, $T_A > T_B$, in (**a**) and (**c**), or defection, $T_A < T_B$, in (**b**) and (**d**). Here, $T_A = 0.5$, and (**a**) $b = 1.2$, , $T_B = 0.4$, $\alpha = 0.4$; (**b**) $b = 2$, $T_B = 0.7$, $\alpha = 0.7$; (**c**) $b = 1.2$, $T_B = 0.4$, $c = 0.5$; (**d**) $b = 2$, $T_B = 0.7$, $c = 0.5$.

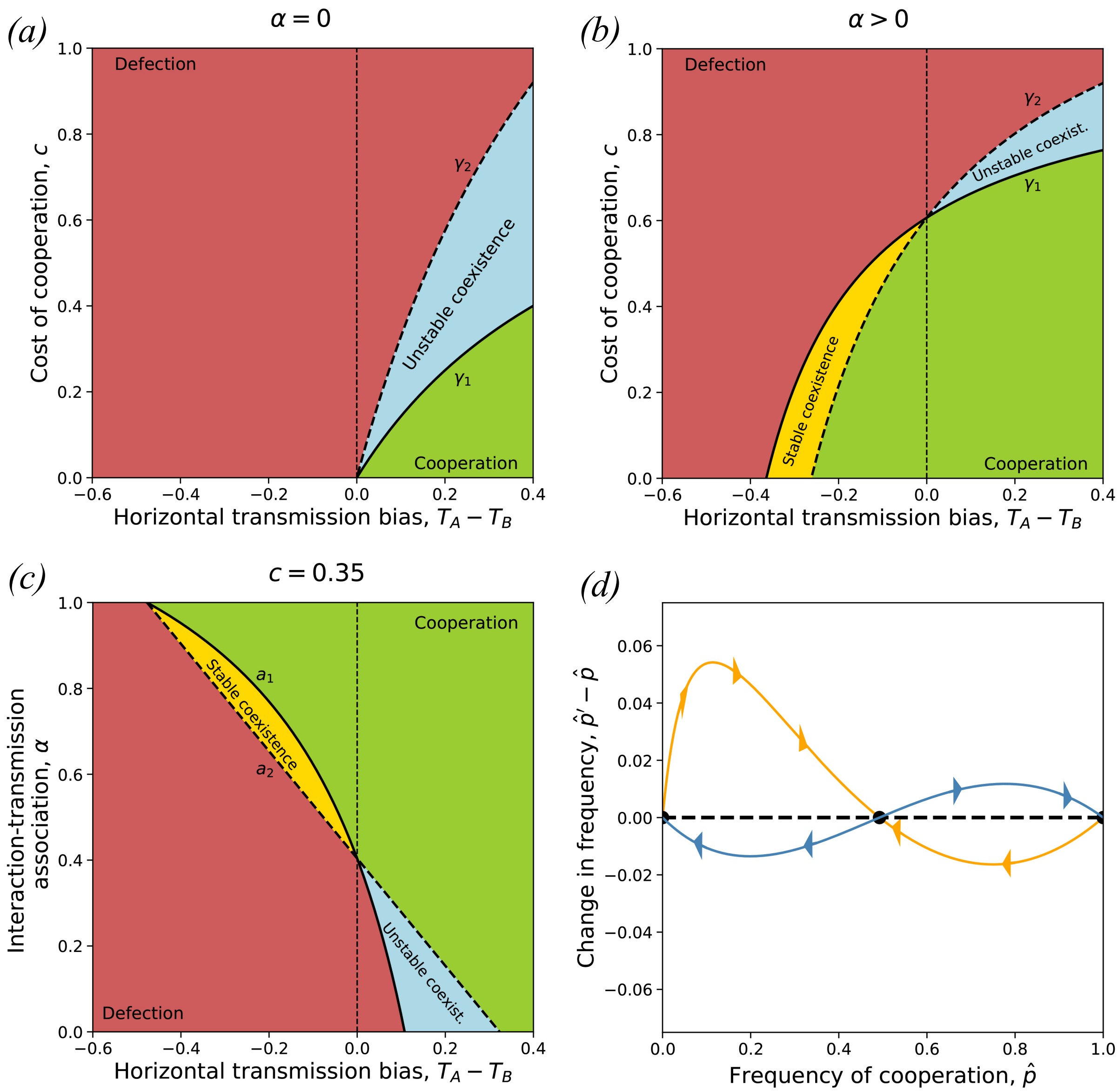

Figure 4: Evolution of cooperation under vertical and horizontal cultural transmission (v=1). The figure shows parameter ranges for global fixation of cooperation (green), global fixation of defection (red), fixation of either cooperation or defection depending on the initial conditions, i.e. unstable polymorphism (blue), and stable polymorphism of cooperation and defection (yellow). (**a-c**) The horizontal transmission bias ($T_A - T_B$) is on the x-axis. In panels **(a)** and **(b)**, the cost of cooperation c is on the y-axis and the cost thresholds γ_1 and γ_2 (Eq. 20) are the solid and dashed lines, respectively. In panel **(c)**, interaction-transmission association α is on the y-axis and the interaction-transmission association thresholds a_1 and a_2 (Eqs. 25) are the solid and dashed lines, respectively. Here, $b = 1.3$, $T_A = 0.4$, $v = 1$, (a) $\alpha = 0$, (b) $\alpha = 0.7$, (c) $c = 0.35$. **(d)** Change in frequency of cooperation among juveniles ($\hat{p}' - \hat{p}$) as a function of the frequency ($\hat{p}$), see Eq. 12. The orange curve shows convergence to a stable polymorphism ($T_A = 0.4$, $T_B = 0.9$, $b = 12$, $c = 0.35$, $v = 1$, and $\alpha = 0.45$). The blue curve shows fixation of either cooperation or defection, depending on the initial frequency ($T_A = 0.5$, $T_B = 0.1$, $b = 1.3$, $c = 0.904$, $v = 1$, and $\alpha = 0.4$). Black circles show the three equilibria.

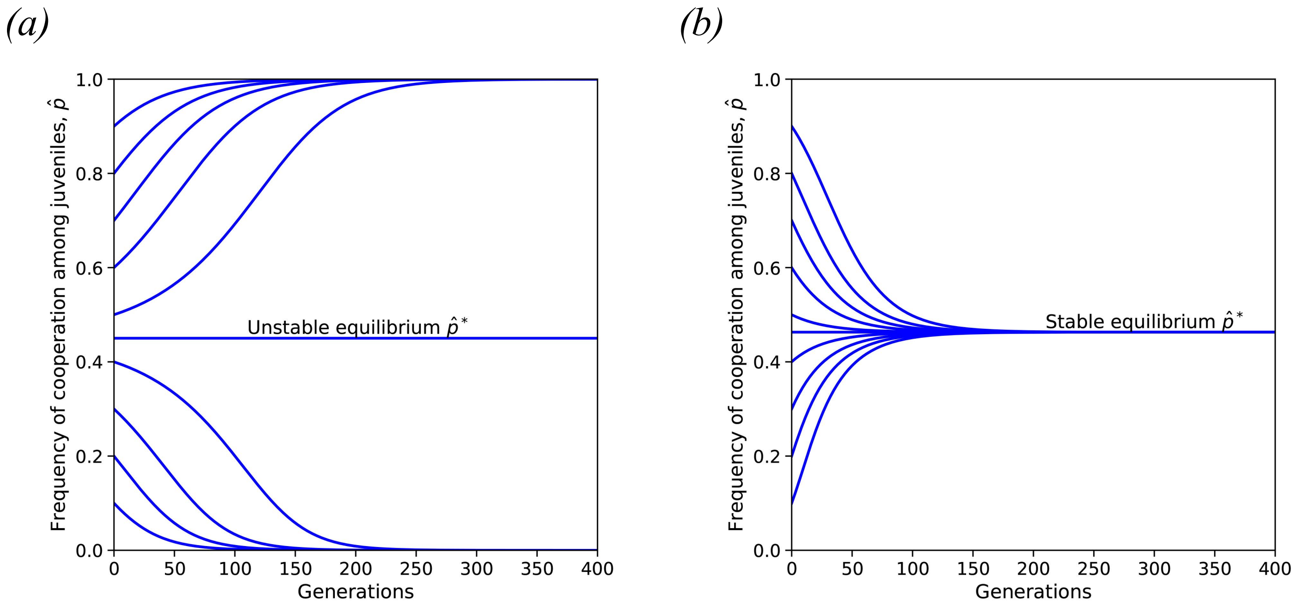

Figure 5: Unstable and stable equilibrium dynamics The figure shows the frequency of cooperation over time (generations) starting from different initial frequencies. In **(a)** the system has unstable equilibrium and therefore, all the curves are moving away from it. In **(b)** the system has stable equilibrium and therefore, all the curves moves towards it. Here, $T_A = 0.5$, $c = 0.5$, $v = 0.59$ and **(a)** $T_B = 0.4$, $\alpha = 0.1$ and $b = 1.2$. **(b)** $T_B = 0.7$, $\alpha = 0.6$ and $b = 2.3$.

3.2 Evolution of interaction-transmission association

We now focus on the evolution of interaction-transmission association, assuming that the population is initially at a stable polymorphism of the two phenotypes, cooperation A and defection B, where the frequency of A among juveniles is $\hat{p}^*$ (Eq. 17). Note that for a stable polymorphism, there must be horizontal bias for defection, $T_A < T_B$, and an intermediate cost of cooperation, $\gamma_2 < c < \gamma_1$ (Eq. 20), see Figure 4b. The equilibrium population mean fitness is $\bar{w}^* = 1 + \hat{p}^*(b - c)$, which is increasing in $\hat{p}^*$, and $\hat{p}^*$ is increasing in α (Appendix B). Therefore, $\bar{w}^*$ increases as α increases. But can this population-level advantage lead to the evolution of increased α?

To answer this question, we add a "modifier locus" [31, 39, 40, 41] that determines the value of α but has no direct effect on fitness. This locus has two alleles, M and m, which induce interaction-transmission associations α_1 and α_2, respectively. Suppose that the population has evolved to a stable equilibrium $\hat{p}^*$ when only allele M is present. We study the local stability of this equilibrium to invasion by the modifier allele m; this is called "external stability" [31, 32].

Denote the frequencies of the pheno-genotypes AM, BM, Am, and Bm by $\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4)$. The frequencies of the pheno-genotypes in the next generation are defined by the recursion system,

$$
\begin{aligned}
\bar{w}\hat{p}_1' =& v\hat{p}_1 x(1+b-c)(1-(1-\alpha_1)(1-x)T_B) + \\
& v\hat{p}_1(1-x)(1-c)(1-\alpha_1 T_B x - T_B(1-x)) + \\
& v\hat{p}_2 x(1+b)T_A(x+\alpha_1(1-x)) + \\
& v\hat{p}_2(1-x)x(1-\alpha_1)T_A + \\
& (1-v)\hat{p}_1(1-(1-x)T_B) + \\
& (1-v)\hat{p}_2 x T_A \ , \\
\bar{w}\hat{p}_2' =& v\hat{p}_1 x(1+b-c)(1-\alpha_1)(1-x)T_B + \\
& v\hat{p}_1(1-x)(1-c)(\alpha_1 T_B + (1-\alpha_1)(1-x)T_B) + \\
& v\hat{p}_2 x(1+b)(1-\alpha_1 T_A(1-x) - T_A x) + \\
& v\hat{p}_2(1-x)(1-(1-\alpha_1)xT_A) + \\
& (1-v)\hat{p}_2(1-xT_A) + \\
& (1-v)\hat{p}_1 T_B(1-x) \ , \\
\bar{w}\hat{p}_3' =& \hat{p}_3 x(1+b-c)(1-(1-\alpha_2)(1-x)T_B) + \\
& \hat{p}_3(1-x)(1-c)(1-\alpha_2 T_B x - T_B(1-x)) + \\
& \hat{p}_4 x(1+b)T_A(x+\alpha_2(1-x)) + \\
& \hat{p}_4(1-x)x(1-\alpha_2)T_A + \\
& (1-v)\hat{p}_3(1-(1-x)T_B) + \\
& (1-v)\hat{p}_4 x T_A \ , \\
\bar{w}\hat{p}_4' =& \hat{p}_3 x(1+b-c)(1-\alpha_2)(1-x)T_B + \\
& \hat{p}_3(1-x)(1-c)(\alpha_2 T_B + (1-\alpha_2)(1-x)T_B) + \\
& \hat{p}_4 x(1+b)(1-\alpha_2 T_A(1-x) - T_A x) + \\
& \hat{p}_4(1-x)(1-(1-\alpha_2)xT_A) + \\
& (1-v)\hat{p}_4(1-xT_A) + \\
& (1-v)\hat{p}_3 T_B(1-x) \ ,
\end{aligned}
\tag{29}
$$

where $x = \hat{p}_1 + \hat{p}_3$ is the total frequency of the cooperative phenotype A, and $\bar{w} = 1 + (b-c)x$ is the population mean fitness.

The stable equilibrium where only allele M is present is $\hat{\mathbf{p}}^* = (\hat{p}^*, 1-\hat{p}^*, 0, 0)$, where

$$
\hat{p}^* = \frac{\alpha b v T_A - cv(1-T_B) + (T_A - T_B)}{\left[c(1-v) - b(1-\alpha_1 v)\right](T_A - T_B)} \ , \tag{30}
$$

setting $\alpha = \alpha_1$ in Eq. 17. $\hat{p}^*$ is a feasible polymorphism ($0 < \hat{p}^* < 1$) if $T_A < T_B$ and $\gamma_2 < c < \gamma_1$ (Result 1).

The local stability of $\hat{\mathbf{p}}^*$ to the introduction of allele m is determined by the linear approximation $\mathbf{L}^*$ of the transformation in Eq. 29 near $\hat{\mathbf{p}}^*$ (i.e. the Jacobian of the transformation at the equilibrium).

$\mathbf{L}^*$ is known to have a block structure, with the diagonal blocks occupied by the matrices $\mathbf{L}^*_{in}$ and $\mathbf{L}^*_{ex}$ [31, 32] . The latter is the external stability matrix: the linear approximation to the transformation near $\hat{\mathbf{p}}^*$ involving only the pheno-genotypes Am and Bm, derived from Eq. 29, with $\bar{w}^* = 1 + (b-c)\hat{p}^*$ as the stable population mean fitness,

$$\mathbf{L}^*_{ex} = \frac{1}{\bar{w}^*}\begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix} = \frac{1}{\bar{w}^*}\begin{bmatrix} \frac{\partial \bar{w}\hat{p}'_3}{\partial \hat{p}_3}(\hat{\mathbf{p}}^*) & \frac{\partial \bar{w}\hat{p}'_3}{\partial \hat{p}_4}(\hat{\mathbf{p}}^*) \\ \frac{\partial \bar{w}\hat{p}'_4}{\partial \hat{p}_3}(\hat{\mathbf{p}}^*) & \frac{\partial \bar{w}\hat{p}'_4}{\partial \hat{p}_4}(\hat{\mathbf{p}}^*) \end{bmatrix} . \tag{31}$$

Because we already assumed that $\mathbf{p}^*$ is internally stable at the beginning of the analysis (i.e. locally stable to small perturbations in the frequencies of AM and BM), the stability of $\mathbf{p}^*$ is determined by the eigenvalues of the external stability matrix $\mathbf{L}^*_{ex}$. This is a positive matrix, and due to the Perron-Frobenius theorem, the leading eigenvalue of $\mathbf{L}^*_{ex}$ is real and positive. Thus, if the leading eigenvalue is less (greater) than one, then the equilibrium $\mathbf{p}^*$ is externally stable (unstable) and allele m cannot (can) invade the population of allele M. The eigenvalues of $\mathbf{L}^*_{ex}$ are the roots of the characteristic polynomial, $R(\lambda)$, which is a quadratic with a positive leading coefficient. Therefore, $\lim_{\lambda\to\pm\infty} R(\lambda) = \infty$, and the leading eigenvalue is less than one (implying stability) if and only if $R(1) > 0$ and $R'(1) > 0$. Thus, a sufficient condition for external instability of $\mathbf{p}^*$ is $R(1) < 0$. $R(\lambda)$ is defined as a determinant, $R(\lambda) = \det(\mathbf{L}^*_{ex} - \lambda\mathbf{I})$, where $\mathbf{I}$ is the 2-by-2 identity matrix. We will use SymPy [38], a Python library for symbolic mathematics to simplify $R(1)$ in the general case when $0 < v \leq 1$ (We did a full analysis without SymPy for the special case when $v = 1$ in Appendix C),

$$\begin{aligned} R(1) = {} & \frac{cv\hat{p}^*[T_A b\hat{p}^{*2}(v\alpha_2 - 1) - 2T_A b\hat{p}^* v\alpha_2 + T_A b\hat{p}^*(1 + v\alpha_2)]}{b\hat{p}(b\hat{p}^* - 2\hat{p}^* + 2) + c\hat{p}^*(c\hat{p}^* - 2) + 1} \\ & + \frac{cv\hat{p}^*[T_A c\hat{p}^{*2}(1 - v) + T_A c\hat{p}^*(v - 1) - T_A\hat{p}^*(1 - c) + T_A]}{b\hat{p}(b\hat{p}^* - 2\hat{p}^* + 2) + c\hat{p}^*(c\hat{p}^* - 2) + 1} \\ & + \frac{cv\hat{p}^*[T_B b\hat{p}^{*2}(1 - v\alpha_2) + T_B b\hat{p}^2(v\alpha_2 - 1) + T_B c\hat{p}^{*2}(v - 1)]}{b\hat{p}(b\hat{p}^* - 2\hat{p}^* + 2) + c\hat{p}^*(c\hat{p}^* - 2) + 1} \\ & + \frac{cv\hat{p}^*[T_B c\hat{p}^*(1 - v) + T_B cv(1 - \hat{p}^* + T_B(\hat{p}^* - 1) + cv(\hat{p}^* - 1))]}{b\hat{p}(b\hat{p}^* - 2\hat{p}^* + 2) + c\hat{p}^*(c\hat{p}^* - 2) + 1} \end{aligned} \tag{32}$$

We should find when $R(1) < 0$. Using SymPy we saw that when $\alpha_1 = \alpha_2$, $R(1) = 0$ which means that there is an eigenvalue that is equal to 1. Now, assume that $\epsilon = \alpha_2 - \alpha_1$. Using the derivative of $R(1)$ with respect to ϵ we will find the sign of $R(1)$ when $\alpha_2 < \alpha_1$.

$$\frac{\partial R(1)}{\partial \epsilon} = \frac{cbv^2\hat{p}^*[(T_A - T_B)\hat{p}^{*2} + (T_B - 2T_A)\hat{p}^* + T_A]}{(\hat{p}^*(b-c) + 1)^2} \tag{33}$$

The denominator is always positive. The numerator is quadratic polynomial of $\hat{p}^*$ with the following roots:

$$\hat{p}^*_1 = \frac{T_A}{T_A - T_B} \quad \hat{p}^*_2 = 1 \tag{34}$$

We assumed $T_B > T_A$ thus, $\hat{p}^*_1 < 0$. The quadratic polynomial has negative leading coefficient $(T_A - T_B)$, and therefore,the numerator is positive for any $\hat{p}^*_1 < \hat{p}^* < 1$. Thus, the derivative of $R(1)$ with respect to ϵ is positive for any ϵ and R(1) grows as $\alpha_2 - \alpha_1$ grows. Therefore, because $R(1)$ grows as $\alpha_2 - \alpha_1$ grows and $R(1) = 0$ when $\alpha_2 - \alpha_1 = 0$, we get that $R(1) < 0$ if and only if $\alpha_2 < \alpha_1$. This is a sufficient condition for external instability. In addition, we also saw that $\frac{\partial R(1)}{\partial \epsilon} > 0$ for every ϵ which makes $\alpha_2 < \alpha_1$ necessary and sufficient condition for successful invasion (external instability).

Result 2. *From a stable polymorphism between cooperation and defection, a modifier allele can successfully invade the population if it decreases the interaction-transmission association α.*

This reduction principle entails that successful invasions will reduce the frequency of cooperation, as well as the population mean fitness (Figure 6). Furthermore, if we a modifier allele that decreases α appears and invades the population from time to time, then the value of α will continue to decrease, further reducing the frequency of cooperation and the population mean fitness. This evolution will proceed as long as there is a stable polymorphism, that is, as long as $a_2 < \alpha < a_1$ (Remark 2, Figure 4c). Thus, we can expect the value of α to approach a_2, the frequency of cooperation to fall to zero, and the population mean fitness to decrease to one (Figure 6).

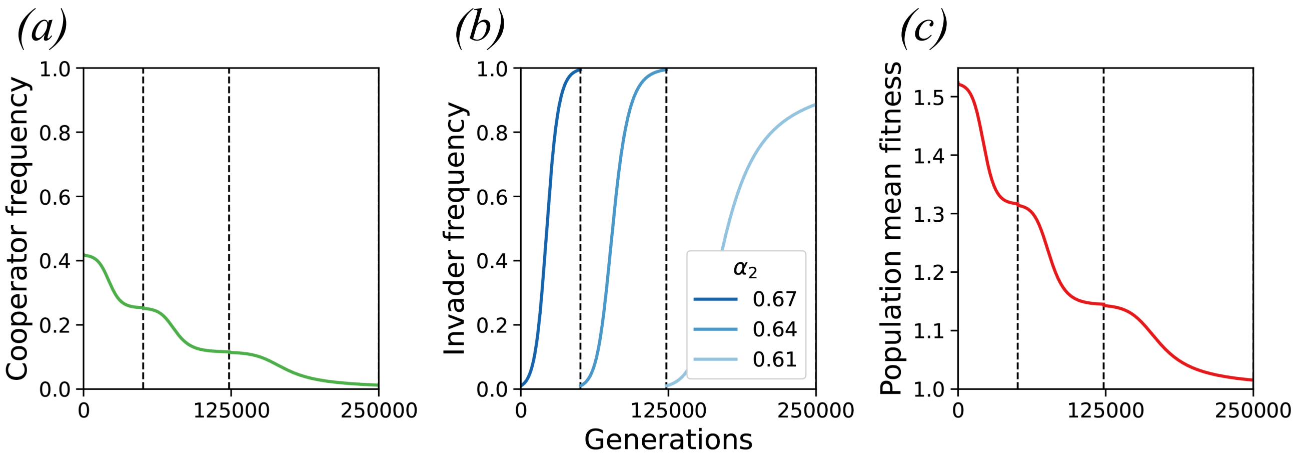

Figure 6: Reduction principle for interaction-transmission association. Consecutive fixation of modifier alleles that reduce interaction-transmission association α in numerical simulations of evolution with two modifier alleles (Eq. D1). When an invading modifier allele is established in the population (frequency > 99.95%), a new modifier allele that reduces interaction-transmission association by 5% is introduced (at initial frequency 0.5%). **(a)** The frequency of the cooperative phenotype A over time. **(b)** The frequency of the invading modifier allele m over time. **(c)** The population mean fitness ($\bar{w}$) over time. Here, $v = 1$, $c = 0.05$, $b = 1.3$, $T_A = 0.4 < T_B = 0.7$, initial interaction-transmission association $\alpha_1 = 0.7$, lower interaction-transmission association threshold $a_2 = 0.605$.

3.3 Population structure

Here, we are going to run stochastic simulations to learn more about evolution of Cooperation when the population is structured. The model is described in subsection 2.3. All the simulations in this section were made by Ohad Lewin-Epstein from Tel Aviv University. The outcomes of stochastic simulations with such a structured population are shown in Figure 7, which demonstrates that the highest cost of cooperation c that permits the evolution of cooperation agrees with the conditions derived above for our model without population structure or stochasticity. An example of stochastic stable polymorphism is shown in Figure 7c. Changing the simulation so that selection is local (i.e. sites can only be settled by offspring of neighboring parents) had only a minor effect on the agreement with the derived conditions (Figure 8).

These comparisons between the deterministic unstructured model and the stochastic structured model show that the conditions derived for the deterministic model can be useful for predicting the dynamics under complex scenarios. Moreover, this structured population model demonstrates that our parameter for interaction-transmission association, α, can represent local interactions between individuals.

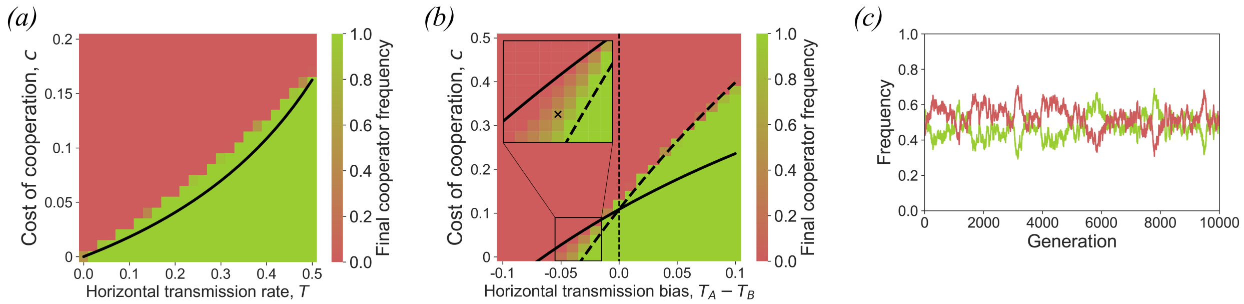

Figure 7: Evolution of cooperation in a structured population. (a-b) The expected frequency of cooperators in a structured population after 10,000 generations is shown (red for 0%, green for 100%) as a function of both the cost of cooperation, c, on the y-axis, and either the symmetric horizontal transmission rate, $T = T_A = T_B$, on the x-axis of panel **(a)**, or the transmission bias, $T_A - T_B$, on the x-axis of panel **(b)**. Black curves represent the cost thresholds for the evolution of cooperation in a well-mixed population with interaction-transmission association, where $\alpha = 1/8$ in inequality 22 for panel **(a)** and in Eqs. 20 for panel **(b)**. The inset in panel **(b)** focuses on an area of the parameter range in which neither phenotype is fixed throughout the simulation, maintaining a stochastic locally stable polymorphism [42]. This stochastic polymorphism is illustrated in panel **(c)**, which shows the frequency of cooperators (green) and defectors (red) over time for the parameter set marked by an x in panel **(b)**. In all cases, the population evolves on a 100-by-100 grid. Cooperation and horizontal transmission are both local between neighbouring sites, and each site has 8 neighbours. Selection operates globally (see Figure S2 for results from a model with local selection). Simulations were stopped at generation 10,000 or if one of the phenotypes fixed. 50 simulations were executed for each parameter set. Benefit of cooperation, $b = 1.3$; perfect vertical transmission $v = 1$. **(a)** Symmetric horizontal transmission, $T = T_A = T_B$; **(b)** Horizontal transmission rate T_A is fixed at 0.4, and T_B varies, $0.3 < T_B < 0.5$. **(c)** Horizontal transmission rates $T_A = 0.4 < T_B = 0.435$ and cost of cooperation $c = 0.02$.

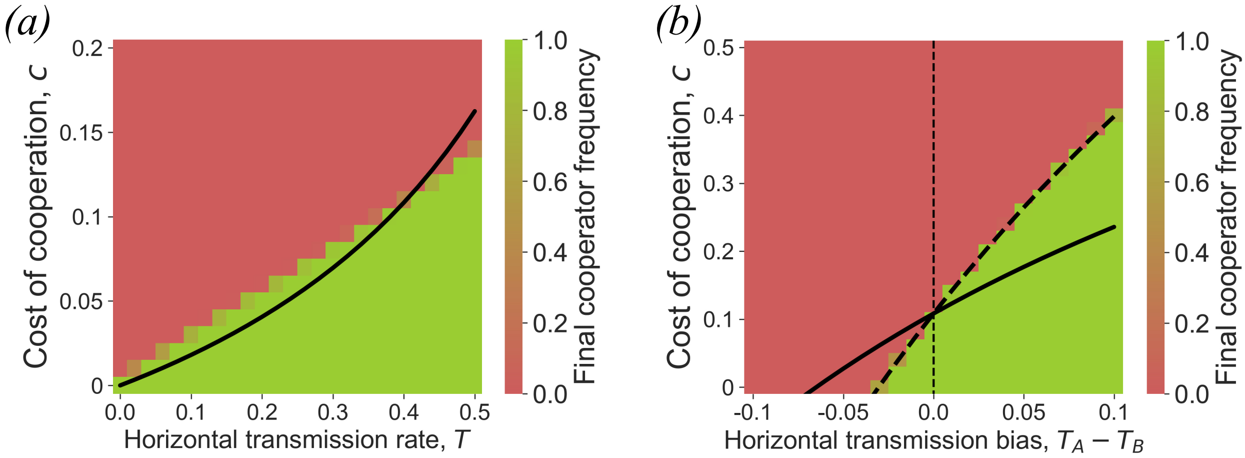

Figure 8: Evolution of cooperation in a structured population with local selection. The expected frequency of cooperators in a structured population after 10,000 generations is shown (red for 0%, green for 100%) as a function of both the cost of cooperation (c) on the y-axis, and the symmetric horizontal transmission rate ($T = T_A = T_B$) on the x-axis of panel **(a)**, or the transmission bias $T_A - T_B$ on the x-axis of panel **(b)**. Cooperation and horizontal transmission are both local between neighbouring sites, and each site had 8 neighbours. Selection operates locally (see Figure 4 for results from a model with global selection). The black curves represent the cost thresholds for the evolution of cooperation in a well-mixed population with interaction-transmission association, where $\alpha = 1/8$ in inequality 14 for panel **(a)** and in Eqs. 12 for panel **(b)**. The population evolves on a 100-by-100 grid. Simulations were stopped at generation 10,000 or if one of the phenotypes fixed. 50 simulations were executed for each parameter set. Here, benefit of cooperation, $b = 1.3$; perfect vertical transmission $v = 1$. **(a)** Symmetric horizontal transmission, $T = T_A = T_B$. **(b)** Horizontal transmission rate T_A is fixed at 0.4, and T_B varies, $0.3 < T_B < 0.5$.

4 Discussion

Under a combination of vertical, oblique, and horizontal transmission with payoffs in the form of a Prisoner's Dilemma game, cooperation or defection can either fix or coexist, depending on the relationship between the cost and benefit of cooperation, the horizontal transmission bias, and the association between social interaction and horizontal transmission (Result 1, Figures 3 and 4). Importantly, cooperation can increase when initially rare (i.e. invade a population of defectors) if and only if, rewriting inequality 21, $c \cdot v(1 - T_B) < b \cdot v\alpha T_A + (T_A - T_B)$, namely, the effective cost of cooperation (left-hand side) is smaller then the effective benefit plus the horizontal transmission bias (right-hand side). This condition cannot be formulated in the form of Hamilton's rule, $c < b \cdot r$, due to the effect of biased horizontal transmission, represented by $(T_A - T_B)$. Remarkably, a polymorphism of cooperation and defection can be stable if horizontal transmission is biased in favor of defection ($T_A < T_B$) and both c and α are intermediate (yellow areas in Figures 3 and 4).

We find that stronger interaction-transmission association α leads to evolution of higher frequency of cooperation and increased population mean fitness. Nevertheless, when cooperation and defection coexist, α is expected to be reduced by natural selection, leading to extinction of cooperation and decreased population mean fitness (Result 2, Figure 6). With $\alpha = 0$, the benefit of cooperation cannot facilitate its evolution; it can only succeed if horizontal transmission is biased in its favor.

Indeed, in our model, horizontal transmission plays a major role in the evolution of cooperation: increasing the transmission of cooperation, T_A, or decreasing the transmission of defection, T_B, facilitates the evolution of cooperation. However, the effect of oblique transmission is more complicated. When there is horizontal transmission bias in favor of cooperation, $T_A > T_B$, increasing the rate of oblique transmission, $1 - v$, will facilitate the evolution of cooperation. In contrast, when the bias is in favor of defection, $T_A < T_B$, higher rates of vertical transmission, v, are advantageous for cooperation, and the rate of vertical transmission must be high enough ($v > \hat{v}$) for cooperation to fix in the population.

Our deterministic model provides a good approximation to outcomes of simulations of a complex stochastic model with population structure in which individuals can only interact with and transmit to their neighbors. In these structured populations interaction-transmission association arises due to both social interactions and horizontal cultural transmission being local (Figure 7 and Figure 8). We did not find any significant difference between local and global selection.

In this work, we have studied a cultural evolution model in which interactions are modeled as a Prisoner's Dilemma game. We also assumed that in each generation individuals have only one meaningful interaction. It would be interesting to investigate a model in which multiple games are played in each generation. This would make the model more realistic and more complicated to solve. To overcome the math complication, numerical simulations can be used. The model can be further expanded by changing the game to Stag Hunt game. Tomasello et al. [36] claim that Stag Hunt game can better explain evolution of cooperation in humans, because humans often collobrate in Stag Hunt type situations in which all participants had alternatives but anticipated an even greater benefit from successful capturing of the stag [43]. Therefore, changing the model so each interaction would be model as Stag Hunt game could help us better understand cooperartion in humans.

In our model we assumed that the model parameters such as vertical transmission(v), horizontal transmission (T_A , T_B), benfit from cooperartion (b), cost of cooperartion (c) and interaction-transmission association (α) do not change over time. However, in reality things are changing. For example, during different time of the year, behavior can be changed due to weather [44], predator abundance [45] and food abundance [46]. This could lead to changes in the model parameters. Further investigation is needed to learn about the effect of such changes in behavior.

Mechanism that was suggested by Traxler and Spichtig [47] showed that *conditional cooperation* based on norm-dependent relational utilities, i.e. individual will only cooperate if it knows that others will cooperate too, can sustain cooperation in a community – provided that cooperation is already at a high frequency. Unlike *conditional cooperation*, interaction-transmission association can sustain cooperation in a community even if cooperation is initially rare. Morsky and Akcay [48] suggested that false beliefs on the frequencies of the cooperator can affect the individual decision whether to cooperate or not. If the individual over-estimates the number of cooperators it will be more likely to cooperate. This false belief can help sustain cooperation even if cooperartion is initially rare.

Feldman et al. [22] studied the dynamics of an altruistic phenotype with vertical cultural transmission and a gene that modifies the transmission of the phenotype. Their results are very sensitive to this genetic modification: without it, the conditions for invasion of the altruistic phenotype reduce to Hamilton's rule. Further work is needed to incorporate such genetic modification of cultural transmission into our model. Woodcock [23] stressed the significance of non-vertical transmission for the evolution of cooperation and carried out simulations with Prisoner's Dilemma payoffs but without horizontal transmission or interaction-transmission association ($\alpha = 0$). Nevertheless, his results demonstrated that it is possible to sustain altruistic behavior via cultural transmission for a substantial length of time. He further hypothesized that horizontal transmission can play an important role in the evolution of cooperation, and our results provide strong evidence for this hypothesis.

To understand the role of horizontal transmission, we first review the role of *assortment*. Eshel and Cavalli-Sforza [18] showed that altruism can evolve when the tendency for *assortative meeting*, i.e. for individuals to interact with others of their own phenotype, is strong enough. Fletcher and Doebeli [19] further argued that a general explanation for the evolution of altruism is given by *assortment*: the correlation between individuals that carry an altruistic trait and the amount of altruistic behavior in their interaction group (see also Bijma and Aanen [20]). They suggested that to explain the evolution of altruism, we should seek mechanisms that generate assortment, such as population structure, repeated interactions, and individual recognition. Our results highlight another mechanism for generating assortment: an association between social interactions and horizontal transmission that creates a correlation between one's partner for interaction and the partner for transmission. This mechanism does not require repeated interactions, population structure, or individual recognition. We show that high levels of such interaction-transmission association greatly increase the potential for evolution of cooperation. With enough interaction-transmission association, cooperation can increase in frequency when initially rare even when there is horizontal transmission bias against it ($T_A < T_B$).

How does non-vertical transmission generate assortment? Lewin-Epstein et al. [24] and Lewin-Epstein and Hadany [25] suggested that microbes that induce their hosts to act altruistically can be favored by selection, which may help to explain the evolution of cooperation. Indeed, it has been shown that microbes can mediate behavioral changes in their hosts [49, 50]. Therefore, natural selection on microbes may favor manipulation of the host so that it cooperates with others. From the kin selection point-of-view, if microbes can be transmitted *horizontally* from one host to another during host interactions, then following horizontal transmission the recipient host will carry microbes that are closely related to those of the donor host, even when the two hosts are (genetically) unrelated. From the assortment point-of-view, infection by behavior-determining microbes during interactions effectively generates assortment because a recipient of help may be infected by a behavior-determining microbe and consequently become a helper. Cultural horizontal transmission can similarly generate assortment between cooperators and enhance the benefit of cooperation if cultural transmission and helping interactions occur between the same individuals, i.e. when there is interaction-transmission association, so that the recipient of help may also be the recipient of the cultural trait for cooperation. Thus, with horizontal transmission, "assortment between focal cooperative players and cooperative acts in their interaction environment" [19] is generated not because the helper is likely to be helped, but rather because the helped is likely to become a helper.

References

[1] Robert Axelrod and William D Hamilton. The evolution of cooperation. *Science*, 211(4489): 1390–1396, 1981.

[2] Lee Alan Dugatkin. *Cooperation among Animals: An Evolutionary Perspective*. Oxford University Press on Demand, 1997.

[3] Adrian V Jaeggi and Michael Gurven. Natural cooperators: food sharing in humans and other primates. *Evolutionary Anthropology: Issues, News, and Reviews*, 22(4):186–195, 2013.

[4] George E Rice and Priscilla Gainer. "Altruism" in the albino rat. *Journal of Comparative and Physiological Psychology*, 55(1):123, 1962.

[5] Peter B Stacey and Walter D Koenig, editors. *Cooperative breeding in birds: long term studies of ecology and behaviour*. Cambridge University Press, 1990.

[6] Indrikis Krams, Tatjana Krama, Kristine Igaune, and Raivo Mänd. Experimental evidence of reciprocal altruism in the pied flycatcher. *Behavioral Ecology and Sociobiology*, 62(4):599–605, 2008.

[7] Barry Sinervo, Alexis Chaine, Jean Clobert, Ryan Calsbeek, Lisa Hazard, Lesley Lancaster, Andrew G McAdam, Suzanne Alonzo, Gwynne Corrigan, and Michael E Hochberg. Self-recognition, color signals, and cycles of greenbeard mutualism and altruism. *Proceedings of the National Academy of Sciences*, 103(19):7372–7377, 2006.

[8] J. B. S. Haldane. *The Causes of Evolution*. Longmans, London, 1932.

[9] William D Hamilton. The genetical evolution of social behaviour. ii. *Journal of Theoretical Biology*, 7(1):17–52, 1964.

[10] Edward O Wilson. Kin selection as the key to altruism: its rise and fall. *Social Research*, pages 159–166, 2005.

[11] Kevin R Foster, Tom Wenseleers, and Francis LW Ratnieks. Kin selection is the key to altruism. *Trends in Ecology & Evolution*, 21(2):57–60, 2006.

[12] Qi Su, Joshua Plotkin, et al. Evolution of cooperation with asymmetric social interactions. *arXiv preprint arXiv:2105.01167*, 2021.

[13] Martin A Nowak. Five rules for the evolution of cooperation. *Science*, 314(5805):1560–1563, 2006.

[14] Luigi Luca Cavalli-Sforza and Marcus W Feldman. *Cultural transmission and evolution: A quantitative approach*. Number 16. Princeton University Press, 1981.

[15] Peter J Richerson and Robert Boyd. *Not by Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press, 2008.

[16] Stephen J Lycett and John AJ Gowlett. On questions surrounding the acheulean 'tradition'. *World Archaeology*, 40(3):295–315, 2008.

[17] Yoav Ram, Uri Liberman, and Marcus W Feldman. Evolution of vertical and oblique transmission under fluctuating selection. *Proceedings of the National Academy of Sciences*, 115(6):E1174–E1183, 2018.

[18] Ilan Eshel and Luigi Luca Cavalli-Sforza. Assortment of encounters and evolution of cooperativeness. *Proceedings of the National Academy of Sciences*, 79(4):1331–1335, 1982.

[19] Jeffrey A. Fletcher and Michael Doebeli. A simple and general explanation for the evolution of altruism. *Proc. R. Soc. B Biol. Sci.*, 276(1654):13–19, 2009.

[20] Piter Bijma and Duur K. Aanen. Assortment, Hamilton's rule and multilevel selection. *Proc. R. Soc. B Biol. Sci.*, 277(1682):673–675, 2010.

[21] Hisashi Ohtsuki, Christoph Hauert, Erez Lieberman, and Martin A. Nowak. A simple rule for the evolution of cooperation on graphs and social networks. *Nature*, 441(7092):502–505, 2006.

[22] Marcus W Feldman, Luca L Cavalli-Sforza, and Joel R Peck. Gene-culture coevolution: models

for the evolution of altruism with cultural transmission. *Proceedings of the National Academy of Sciences*, 82(17):5814–5818, 1985.

[23] Scott Woodcock. The significance of non-vertical transmission of phenotype for the evolution of altruism. *Biology and Philosophy*, 21(2):213–234, 2006.

[24] Ohad Lewin-Epstein, Ranit Aharonov, and Lilach Hadany. Microbes can help explain the evolution of host altruism. *Nature Communications*, 8:14040, 2017.

[25] Ohad Lewin-Epstein and Lilach Hadany. Host-microbiome coevolution can promote cooperation in a rock-paper-scissors dynamics. *Proc. R. Soc. B Biol. Sci.*, 287(1920):20192754, feb 2020.

[26] Yael Gurevich, Ohad Lewin-Epstein, and Lilach Hadany. The evolution of paternal care: a role for microbes? *Philos. Trans. R. Soc. B Biol. Sci.*, 375(1808):20190599, sep 2020.

[27] Carla Handley and Sarah Mathew. Human large-scale cooperation as a product of competition between cultural groups. *Nature communications*, 11(1):1–9, 2020.

[28] Martin A Nowak and Karl Sigmund. Tit for tat in heterogeneous populations. *Nature*, 355(6357): 250–253, 1992.

[29] Jerome M Sacks. A stable equilibrium with minimum average fitness. *Genetics*, 56(4):705, 1967.

[30] Sabin Lessard et al. Evolutionary stability: one concept, several meanings. *Theoretical population biology*, 37(1):159–170, 1990.

[31] Uri Liberman and Marcus W. Feldman. Modifiers of mutation rate: A general reduction principle. *Theor. Popul. Biol.*, 30:125–142, 1986.

[32] Lee Altenberg, Uri Liberman, and Marcus W. Feldman. Unified reduction principle for the evolution of mutation, migration, and recombination. *Proc. Natl. Acad. Sci. U. S. A.*, 114(12): E2392–E2400, mar 2017.

[33] Matthew R. Zefferman. Mothers teach daughters because daughters teach granddaughters: the evolution of sex-biased transmission. *Behav. Ecol.*, 27(4):1172–1181, 2016.

[34] Manfred Milinski. Tit for tat in sticklebacks and the evolution of cooperation. *nature*, 325(6103): 433–435, 1987.

[35] Brian Skyrms. *The stag hunt and the evolution of social structure*. Cambridge University Press, 2004.

[36] Michael Tomasello, Alicia P Melis, Claudio Tennie, Emily Wyman, Esther Herrmann, Ian C Gilby, Kristen Hawkes, Kim Sterelny, Emily Wyman, Michael Tomasello, et al. Two key steps in the evolution of human cooperation: The interdependence hypothesis. *Current anthropology*, 53(6):000–000, 2012.

[37] Edward F Moore. Machine models of self-reproduction. In *Proceedings of symposia in applied mathematics*, volume 14, pages 17–33. American Mathematical Society New York, 1962.

[38] Aaron Meurer, Christopher P Smith, Mateusz Paprocki, Ondřej Čertík, Sergey B Kirpichev, Matthew Rocklin, AMiT Kumar, Sergiu Ivanov, Jason K Moore, Sartaj Singh, et al. Sympy: symbolic computing in python. *PeerJ Computer Science*, 3:e103, 2017.

[39] Marcus W. Feldman. Selection for linkage modification: I. Random mating populations. *Theor. Popul. Biol.*, 3:324–346, 1972.

[40] Uri Liberman and Marcus W. Feldman. A general reduction principle for genetic modifiers of recombination. *Theor. Popul. Biol.*, 30(3):341–71, dec 1986.

[41] Uri Liberman. External stability and ESS: criteria for initial increase of new mutant allele. *J. Math. Biol.*, 26(4):477–485, 1988.

[42] Samuel Karlin, Uri Lieberman, and Uri Liberman. Random temporal variation in selection intensities: One-locus two-allele model. *J. Math. Biol.*, 6(3):1–17, 1975.

[43] John C Mitani, Josep Call, Peter M Kappeler, Ryne A Palombit, and Joan B Silk. *The evolution*

of primate societies. University of Chicago Press, 2012.

[44] Francisco M Azcárate, Eva Kovacs, and Begoña Peco. Microclimatic conditions regulate surface activity in harvester ants messor barbarus. *Journal of Insect Behavior*, 20(3):315–329, 2007.

[45] Mark Briffa, Simon D Rundle, and Adam Fryer. Comparing the strength of behavioural plasticity and consistency across situations: animal personalities in the hermit crab pagurus bernhardus. *Proceedings of the Royal Society B: Biological Sciences*, 275(1640):1305–1311, 2008.

[46] Chris J Johnson, Katherine L Parker, and Douglas C Heard. Foraging across a variable landscape: behavioral decisions made by woodland caribou at multiple spatial scales. *Oecologia*, 127(4): 590–602, 2001.

[47] Christian Traxler and Mathias Spichtig. Social norms and the indirect evolution of conditional cooperation. *Journal of Economics*, 102(3):237–262, 2011.

[48] Bryce Morsky and Erol Akcay. False beliefs can bootstrap cooperative communities through social norms. 2020.

[49] Andrew P Dobson. The population biology of parasite-induced changes in host behavior. *The Quarterly Review of Biology*, 63(2):139–165, 1988.

[50] Robert Poulin. Parasite manipulation of host behavior: an update and frequently asked questions. In *Advances in the Study of Behavior*, volume 41, pages 151–186. Elsevier, 2010.

Appendices

Appendix A Local stability criterion

Let $f(p) = \lambda \cdot (p' - p)$, where $\lambda > 0$, and 0 and 1 are equilibria, that is, $f(0) = 0$ and $f(1) = 0$.

Set $p > p^* = 0$. Using a linear approximation for $f(p)$ near 0, we have

$$p' < p \Leftrightarrow f(p)/p < 0 \Leftrightarrow \frac{f'(0) \cdot p + O(p^2)}{p} < 0 \Leftrightarrow f'(0) + O(p) < 0 \;. \tag{A1}$$

Therefore, by definition of big-O notation, if $f'(0) < 0$ then there exists $\epsilon > 0$ such that for any local perturbation $0 < p < \epsilon$, it is guaranteed that $0 < p' < p$; that is, p' is closer to zero than p.

Set $p < p^* = 1$ Using a linear approximation for $f(p)$ near 1, we have

$$1 - p' < 1 - p \Leftrightarrow -\frac{f(p)}{1-p} < 0 \Leftrightarrow \frac{f'(1)(p-1) + O\big((p-1)^2\big)}{p-1} < 0 \Leftrightarrow f'(1) - O(1-p) < 0 \;. \tag{A2}$$

Therefore, if $f'(1) < 0$ then there exists $\epsilon > 0$ such that for any $1-\epsilon < 1-p < 1$ we have $1-p' < 1-p$; that is, p' is closer to one than p.

Appendix B Effect of interaction-transmission association on mean fitness

To determine the effect of increasing α on the stable population mean fitness, $\bar{w}^* = 1 + (b-c)\hat{p}^*$, we must analyze its effect on $\hat{p}^*$,

$$\frac{\partial \hat{p}^*}{\partial \alpha} = \frac{bT_A - c(1-T_B) + (T_A - T_B)}{b(1-\alpha)^2(T_B - T_A)} \;. \tag{B1}$$

Note that stable polymorphism implies $c < \gamma_1$, and because $\alpha < 1$, we have

$$c < \gamma_1 = \frac{b\alpha T_A + (T_A - T_B)}{1 - T_B} < \frac{bT_A + (T_A - T_B)}{1 - T_B}. \tag{B2}$$

Therefore, the numerator in Eq. B1 is positive. Since $T_A < T_B$, the denominator in Eq. B1 is also positive, and hence the derivative $\partial \hat{p}^*/\partial \alpha$ is positive. Thus, the population mean fitness increases as interaction-transmission association α increases.

Appendix C Reduction principle

Here, we assume perfect vertical transmission $v = 1$. We start from Eq. 31 and we substitute $v = 1$ and we get

$$\mathbf{L}^*_{ex} = \frac{1}{\bar{w}^*}\begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix} = \frac{1}{\bar{w}^*}\begin{bmatrix} \frac{\partial \bar{w}\hat{p}'_3}{\partial \hat{p}_3}(\hat{\mathbf{p}}^*) & \frac{\partial \bar{w}\hat{p}'_3}{\partial \hat{p}_4}(\hat{\mathbf{p}}^*) \\ \frac{\partial \bar{w}\hat{p}'_4}{\partial \hat{p}_3}(\hat{\mathbf{p}}^*) & \frac{\partial \bar{w}\hat{p}'_4}{\partial \hat{p}_4}(\hat{\mathbf{p}}^*) \end{bmatrix} =$$
$$\frac{1}{\bar{w}^*}\begin{bmatrix} (1 + b\dot{p}^* - c)(1 - T_B(1 - \dot{p}^*)) + b\dot{p}^*\alpha_2 T_B(1 - \dot{p}^*) & (1 + b\dot{p}^*)T_A\dot{p}^* + b\dot{p}^*\alpha_2 T_A(1 - \dot{p}^*) \\ (1 + b\dot{p}^* - c)T_B(1 - \dot{p}^*) - b\dot{p}^*\alpha_2 T_B(1 - \dot{p}^*) & (1 + b\dot{p}^*)(1 - T_A\dot{p}^*) - b\dot{p}^*\alpha_2 T_A(1 - \dot{p}^*) \end{bmatrix} \tag{C1}$$

Since multiplication by a positive factor doesn't change the sign, and using the properties of the determinant, we have

$$\begin{aligned} \operatorname{sign} R(1) &= \operatorname{sign}\det(\mathbf{L}^*_{ex} - \mathbf{I}) = \operatorname{sign}(\bar{w}^*)^2 \det(\mathbf{L}^*_{ex} - \mathbf{I}) = \\ &\operatorname{sign}\det(\bar{w}^*\mathbf{L}^*_{ex} - \bar{w}^*\mathbf{I}) = \operatorname{sign}\det\begin{bmatrix} l_{11} - \bar{w}^* & l_{12} \\ l_{21} & l_{22} - \bar{w}^* \end{bmatrix}, \end{aligned} \tag{C2}$$

where l_{ij} are defined in Eq. 31. Adding the second row in Eq. C2 to the first row, which does not change the determinant, and substituting $\bar{w}^* = 1 + (b - c)\dot{p}^*$, we get

$$\begin{aligned} \operatorname{sign} R(1) &= \operatorname{sign}\det\begin{bmatrix} -c(1 - \dot{p}^*) & c\dot{p}^* \\ (1 - \dot{p}^*)\big[(1 + b\dot{p}^* - c)T_B - b\alpha_2 T_B\dot{p}^*\big] & \dot{p}^*\big[-(1 + b\dot{p}^*)T_A - b\alpha_2 T_A(1 - \dot{p}^*) + c\big] \end{bmatrix} = \\ &\operatorname{sign}\left[c\dot{p}^*(1 - \dot{p}^*)\cdot\det\begin{bmatrix} -1 & 1 \\ (1 + b\dot{p}^* - c)T_B - b\alpha_2 T_B\dot{p}^* & -(1 + b\dot{p}^*)T_A - b\alpha_2 T_A(1 - \dot{p}^*) + c \end{bmatrix}\right] = \\ &\operatorname{sign}\det\begin{bmatrix} -1 & 1 \\ (1 + b\dot{p}^* - c)T_B - b\alpha_2 T_B\dot{p}^* & -(1 + b\dot{p}^*)T_A - b\alpha_2 T_A(1 - \dot{p}^*) + c \end{bmatrix}, \end{aligned} \tag{C3}$$

since $c > 0$, $0 < \dot{p}^* < 1$. That is,

$$\begin{aligned} \operatorname{sign} R(1) &= \operatorname{sign}\Big[(1 + b\dot{p}^*)T_A + b\alpha_2 T_A(1 - \dot{p}^*) - c - (1 + b\dot{p}^* - c)T_B + b\dot{p}^*\alpha_2 T_B\Big] = \\ &\operatorname{sign}\Big[(1 + b(1 - \alpha_2)\dot{p}^*)(T_A - T_B) + b\alpha_2 T_A - c(1 - T_B)\Big]. \end{aligned} \tag{C4}$$

Substituting $\dot{p}^*$ from Eq. 30, we get

$$\begin{aligned} R(1) < 0 \Leftrightarrow &\big[c(1 - T_B) - b\alpha_1 T_A - (T_A - T_B)\big]\frac{1 - \alpha_2}{1 - \alpha_1} - c(1 - T_B) + b\alpha_2 T_A + (T_A - T_B) < 0 \Leftrightarrow \\ &(1 - \alpha_2)\big[c(1 - T_B) - b\alpha_1 T_A - (T_A - T_B)\big] < (1 - \alpha_1)\big[c(1 - T_B) - b\alpha_2 T_A - (T_A - T_B)\big] \Leftrightarrow \\ &- b\alpha_1 T_A - \alpha_2 c(1 - T_B) + \alpha_2(T_A - T_B) < -b\alpha_2 T_A - \alpha_1 c(1 - T_B) + \alpha_1(T_A - T_B) \Leftrightarrow \\ &\alpha_1\big[c(1 - T_B) - bT_A - (T_A - T_B)\big] < \alpha_2\big[c(1 - T_B) - bT_A - (T_A - T_B)\big] \Leftrightarrow \\ &\alpha_1\big[bT_A + (T_A - T_B) - c(1 - T_B)\big] > \alpha_2\big[bT_A + (T_A - T_B) - c(1 - T_B)\big]. \end{aligned} \tag{C5}$$

We assumed $c < \gamma_1$, and since $0 \leq \alpha_1 \leq 1$,

$$\begin{aligned} & c < \gamma_1 = \frac{b\alpha_1 T_A + (T_A - T_B)}{1 - T_B} \iff \\ & 0 < b\alpha_1 T_A + (T_A - T_B) - c(1 - T_B) \implies \\ & 0 < bT_A + (T_A - T_B) - c(1 - T_B) \ . \end{aligned} \tag{C6}$$

Combining inequalities C5 and C6, we find that $R(1) < 0$ if and only if $\alpha_1 > \alpha_2$, which is a sufficient condition for external instability. Therefore, if α_2, the interaction-transmission association of the invading modifier allele m, is less than α_1, the interaction-transmission association of the resident allele M, then invasion will be successful.

Determining a necessary and sufficient condition for successful invasion is more complicated, requiring analysis of the sign of $R'(1)$. We did it in the general case (starting from Eq. 33).