

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

immediate

4

June 8, 2021

This work was carried out under the supervision of **Dr. Yoav Ram** from the Efi Arazi  
6 School of Computer Science, The Interdisciplinary Center, Herzliya.

# Abstract

8 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  
10 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  
12 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  
14 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  
16 with Hamilton's rule incorporating relatedness and assortment.

# Contents

18	<b>1 Introduction</b>	<b>5</b>
	<b>2 Related Work</b>	<b>7</b>
20	<b>3 Models</b>	<b>9</b>
	<b>4 Results</b>	<b>12</b>
22	4.1 Evolution of cooperation . . . . .	12
	4.2 Evolution of interaction-transmission association . . . . .	18
24	4.3 Population structure . . . . .	20
	<b>5 Discussion</b>	<b>23</b>
26	<b>A Local stability criterion</b>	<b>28</b>
	<b>B Equilibria and stability</b>	<b>29</b>
28	<b>C Effect of interaction-transmission association on mean fitness</b>	<b>31</b>
	<b>D Reduction principle</b>	<b>32</b>

## 30 1 Introduction

Cooperative behavior can reduce an individual's fitness and increase the fitness of its conspecifics or  
32 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  
34 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  
36 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  
38 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  
40 reproductive cost to the actor that cooperates,  $c$ , is less than the benefit to the recipient,  $b$ , times the  
relatedness,  $r$ , between the recipient and the actor. This condition is known as *Hamilton's rule*:

$$42 \quad c < b \cdot r, \quad (1)$$

where the relatedness coefficient  $r$  measures the probability that an allele sampled from the cooperator  
44 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  
46 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  
48 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  
50 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  
52 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  
54 a high benefit: cost ratio will favour cooperation.

Beside kin selection, two more major theories were suggested to explain to evolution of coopera-  
56 tion.

**Reciprocity** suggests repeating interactions or individual recognition as key factors for explaining the  
58 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  
60 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  
62 evolution of cooperation if the cost is smaller than  $w$  the probability for another encounter between  
the same two individuals multiplied by the benefit.

$$64 \quad c < bw \quad (2)$$

Direct reciprocity assumes that both players are in a position to cooperate. Direct reciprocity can  
66 not explain cooperation in asymmetric interactions. In humans, such interactions happen often, for

example humans often donate money. *Indirect reciprocity* has been suggested to explain this behavior.

68 Nowak [12] claims that direct reciprocity is like a barter economy based on the immediate exchange of  
goods, while indirect reciprocity resembles the invention of money. The money that "fuels the engines"  
70 of indirect reciprocity is reputation. However, Reciprocity assume repeating interactions and therefore,  
has difficulty explaining evolution of cooperation if the no repeating interactions occurs.

72 **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  
74 selection work as is: the population is divided into groups. In each groups there are cooperators, which  
help to other group members and defectors which do not help. Individuals reproduce proportional to  
76 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 grow faster will split more often. Groups of cooperators are growing faster than  
78 group of defectors. Therefore, cooperation can evolve in this model when the ratio between benefit  
b and cost c is more than one plus the ratio between the maximum group size n and the number of  
80 groups m:

$$\frac{b}{c} > 1 + \frac{n}{m} \quad (3)$$

82 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.  
84 Cooperative behavior can 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  
86 other modes of communication [13, 14]. Cultural transmission may be modeled as vertical, horizontal,  
or oblique: vertical transmission occurs between parents and offspring, horizontal transmission occurs  
88 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  
90 transmission models can be be more rapid than under pure vertical transmission [13, 15, 16].

Here, we study models for the cultural evolution of cooperation that include both vertical and non-  
92 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  
94 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  
96 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  
98 cannot, partly because it facilitates the generation of assortment [17], and partly because it diminishes  
the effect of selection (due to non-vertical transmission from non-reproducing individuals [16]).

## 100 2 Related Work

102 Eshel and Cavalli-Sforza [18] studied a related model for the evolution of cooperative behavior. Their  
 104 model included *assortative meeting*, or non-random encounters, where a fraction  $m$  of individuals in  
 106 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]

$$108 c < b \cdot m, \quad (4)$$

where  $b$  and  $c$  are the benefit and cost of cooperation<sup>1</sup>.

The role of assortment in the evolution of altruism was emphasized by Fletcher and Doebeli [17].  
 110 They found that in a *public-goods* game, altruism will evolve if cooperative individuals experience  
 112 more cooperation, on average, than defecting individuals, and “thus, the evolution of altruism requires  
 (positive) assortment between focal *cooperative* players and cooperative acts in their interaction  
 environment.” With some change in parameters, this condition is summarized by [17, eq. 2.3]

$$114 c < b \cdot (p_C - p_D), \quad (5)$$

where  $p_C$  is the probability that a cooperator receives help, and  $p_D$  is the probability that a defector  
 116 receives help<sup>2</sup>. Bijma and Aanen [19] obtained a result related to inequality 5 for other games.

Cooperation can also evolve when interactions are determined by population structure. For example,  
 118 Ohtsuki et al. [20] 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  
 120 much larger than  $k$  (i.e. sparse structure), they found that cooperative behaviour can evolve if [20]

$$c < b \cdot \frac{1}{k}. \quad (6)$$

122 They thus interpret  $1/k$  as *social relatedness* or *social viscosity* [20].

Feldman et al. [21] introduced the first model for the evolution of altruism by cultural transmission  
 124 with kin selection and demonstrated that if the fidelity of cultural transmission of altruism is  $\varphi$ , then  
 the condition for evolution of altruism in the case of sib-to-sib altruism is [21, Eq. 16]

$$126 c < b \cdot \varphi - \frac{1 - \varphi}{\varphi}. \quad (7)$$

In inequality 7,  $\varphi$  replaces relatedness ( $r$  in inequality 1) or assortment ( $m$  in inequality 4), but the ef-  
 128 fective 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).

---

<sup>1</sup>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].

<sup>2</sup>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.

130 Both Woodcock [22] and Lewin-Epstein et al. [23] demonstrated that non-vertical transmission can help  
131 explain the evolution of cooperative behavior, the former using simulations with cultural transmission,  
132 the latter using a model where cooperation is mediated by host-associated microbes. Indeed, models  
133 in which microbes affect their host's behavior [23, 24, 25] are mathematically similar to models of  
134 cultural transmission, and they also emphasize the role of non-vertical transmission [13].

### 3 Models

136 Consider a large population whose members can be one of two phenotypes:  $\phi = A$  for cooperators  
 or  $\phi = B$  for defectors. An offspring inherits its phenotype from its parent via vertical transmission  
 138 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. [16], given that the parent's phenotype is  
 140  $\phi$  and assuming uni-parental inheritance [26], the conditional probability that the phenotype  $\phi'$  of the  
 offspring is  $A$  is

$$142 \quad P(\phi' = A | \phi) = \begin{cases} v + (1 - v)p, & \text{if } \phi = A \\ (1 - v)p, & \text{if } \phi = B \end{cases}, \quad (8)$$

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

144 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  
 146 transmission) is

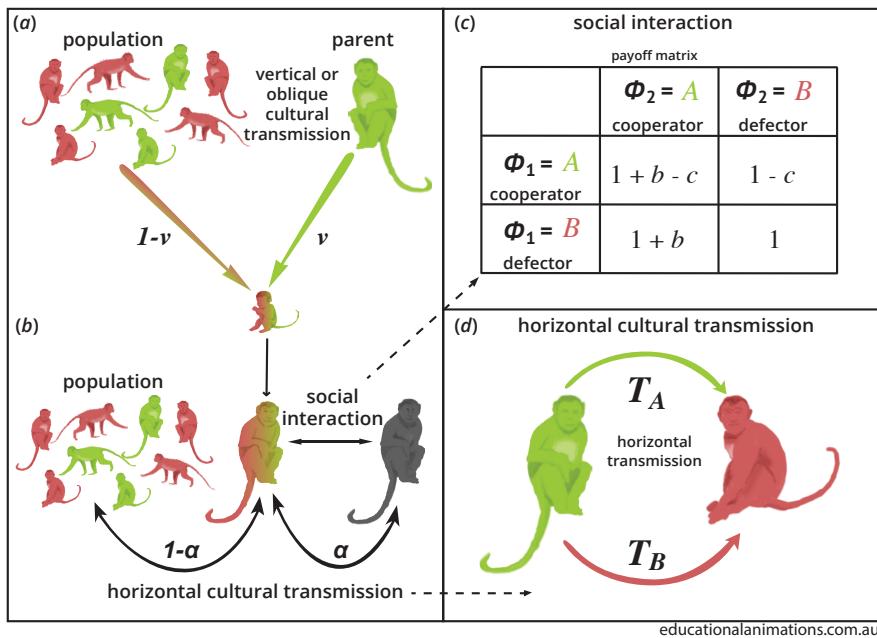
$$\hat{p} = \dot{p}[v + (1 - v)p] + (1 - \dot{p})[(1 - v)p] = v\dot{p} + (1 - v)p. \quad (9)$$

148 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  
 150 assume  $c < b$ . Figure 1a shows the payoff matrix, i.e. the fitness of an individual with phenotype  $\phi_1$   
 when interacting with a partner of phenotype  $\phi_2$ . The choice of prisoner's dilemma as the interaction  
 152 model was motivated by the fact that prisoner's dilemma is a common game used to study evolution  
 of cooperation[1], [27], [28]. Although we decided to focus on prisoner's dilemma, other games such  
 154 as stag hunt[29] may be a better explanation of cooperation behavior in humans[30].

Social interactions occur randomly: two juvenile individuals with phenotype  $A$  interact with proba-  
 156 bility  $\hat{p}^2$ , 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  
 158 between pairs of individuals from the same generation. It occurs between socially interacting partners  
 with probability  $\alpha$ , or between a random pair with probability  $1 - \alpha$  (see Figure 1b). However,  
 160 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  
 162 (Table 1, Figure 1d). Thus, the frequency  $p'$  of phenotype  $A$  among adults in the next generation, after  
 horizontal transmission, is

$$164 \quad p' = \hat{p}^2[\alpha + (1 - \alpha)(\hat{p} + (1 - \hat{p})(1 - T_B))] + \\ \hat{p}(1 - \hat{p})[\alpha(1 - T_B) + (1 - \alpha)(\hat{p} + (1 - \hat{p})(1 - T_B))] + \\ (1 - \hat{p})\hat{p}[\alpha T_A + (1 - \alpha)\hat{p}T_A] + (1 - \hat{p})^2[(1 - \alpha)\hat{p}T_A] \\ = \hat{p}^2(T_B - T_A) + \hat{p}(1 + T_A - T_B). \quad (10)$$

The frequency of  $A$  among parents (i.e. after selection) follows a similar dynamic, but also includes



**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  $\alpha$ , where  $\alpha$  is the interaction-transmission association parameter. **(c)** The prisoner's dilemma payoff matrix shows the fitness of phenotype  $\phi_1$  when interacting with phenotype  $\phi_2$ . **(d)** The probabilities of successful horizontal cultural transmission of phenotypes  $A$  (cooperator) and  $B$  (defector) are  $T_A$  and  $T_B$ , respectively.

166 the effect of natural selection, and is therefore

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

168 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  $\dot{p}'$   
170 from Eq. 11 and obtain

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

172 Table 2 lists the model variables and parameters.

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

Phenotype $\phi_1$	Phenotype $\phi_2$	Frequency	Fitness of $\phi_1$	$P(\phi_1 = A)$ via horizontal transmission:	
				from partner, $\alpha$	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	$c < b$
$\alpha$	Probability of interaction-transmission association	$[0, 1]$
$T_A, T_B$	Horizontal transmission rates of phenotype $A$ and $B$	$(0, 1)$

## 4 Results

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

### 4.1 Evolution of cooperation

177 To learn about the evolution of cooperation we investigate local stability of the equilibria of the model  
 178 in Eq. 12. The equilibria are the solutions of  $\dot{p}' - \dot{p} = 0$ . Note that Eq. 12 is not simple polynomial  
 179 but a polynomial fraction, therefore the solution is not trivial. The fixed points (equilibria) of the  
 180 recursion (Eq. 12) are  $\hat{p} = 0$ ,  $\hat{p} = 1$ , and (see Appendix B Eq. B5)

$$182 \quad \hat{p}^* = \frac{\alpha bvT_A - cv(1 - T_B) + (T_A - T_B)}{[c(1 - v) - b(1 - \alpha v)](T_A - T_B)} . \quad (13)$$

Define the following cost thresholds,  $\gamma_1$  and  $\gamma_2$ , and the vertical transmission threshold,  $\hat{v}$ ,

$$184 \quad \gamma_1 = \frac{bvaT_A + (T_A - T_B)}{v(1 - T_B)}, \quad \gamma_2 = \frac{bvaT_B + (1 + b)(T_A - T_B)}{v(1 - T_B) + (1 - v)(T_A - T_B)}, \quad \hat{v} = \frac{T_B - T_A}{1 - T_A} . \quad (14)$$

Then we have the following result.

186

**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  $\gamma_1$  and  $\gamma_2$  and the vertical transmission threshold  $\hat{v}$  (Eq. 14) :*

190 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  $\gamma_1 < c$ .

192 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  
 $\gamma_2 < c < \gamma_1$ .

194 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  
196 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  
198 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  
200 cooperation and the vertical transmission rate. These conditions are illustrated in Figures 2a, 2b, 3a,  
and 3b, and the analysis is in Appendix B. Note that *stable* and *unstable* polymorphism are also called,  
202 respectively, *coexistence* and *bistable competition*.

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

206 **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,

208 
$$c < \gamma_1 = \frac{bvaT_A + (T_A - T_B)}{v(1 - T_B)} . \quad (15)$$

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

Importantly, increasing interaction-transmission association  $\alpha$  increases the cost threshold ( $\partial\gamma_1/\partial\alpha >$   
212 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  
214 the evolution of cooperation ((Figure 3a and 3b). However, increasing the horizontal transmission of  
defection,  $T_B$ , can increase or decrease the cost threshold, but it increases the cost threshold when  
216 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  
218 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  
220 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,  
222 making it easier for cooperation to increase when rare (Figure 2b). 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,  
224 making it harder for cooperation to increase when rare (Figure 2a).

In general, this condition cannot be formulated in the form of Hamilton's rule due to the bias in  
226 horizontal transmission, represented by  $T_A - T_B$ . If  $T_A = T_B$ , then, from Result 1 and inequality 15,

cooperation will take over the population from any initial frequency if the cost is low enough,

$$228 \quad c < b \cdot \frac{\alpha T}{1 - T}, \quad (16)$$

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 16 equals  $\gamma_1$  when  $T = T_A = T_B$ .

From inequality 15, 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

$$c < \frac{T_A - T_B}{v(1 - T_B)}. \quad (17)$$

236 Figure 3a illustrates this condition (for  $v = 1$ ), which is obtained by setting  $\alpha = 0$  in inequality 15. In this case, the benefit of cooperation,  $b$ , does not affect the evolution of cooperation, and the 238 outcome is determined only by cultural transmission. Further, inequality 15 shows that with perfect interaction-transmission association ( $\alpha = 1$ ), cooperation will increase when rare if

$$240 \quad c < \frac{bvT_A + (T_A - T_B)}{v(1 - T_B)}. \quad (18)$$

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 18 applies (Figure 3). This is similar to case of microbe-induced cooperation studied by Lewin-Epstein et al. 244 [23]; therefore when  $v = 1$ , this remark is equivalent to their eq. 1.

It is interesting to examine the general effect of interaction-transmission association  $\alpha$  on the evolution 246 of cooperation. Define the interaction-transmission association thresholds,  $a_1$  and  $a_2$ , as

$$a_1 = \frac{c \cdot v(1 - T_A) - (T_A - T_B)(1 + b - c)}{b \cdot v \cdot T_B}, \quad a_2 = \frac{c \cdot v(1 - T_B) - (T_A - T_B)}{b \cdot v \cdot T_A}. \quad (19)$$

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

250 Figures 2c and 2d 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 252 bias favoring defection,  $T_A < T_B$ , cooperation can fix from any frequency if  $\alpha$  is large enough,  $a_1 < \alpha$  (green area with  $T_A < T_B$ ), and can reach stable polymorphism if  $\alpha$  is intermediate,  $a_2 < \alpha < a_1$  254 (yellow area). Without horizontal bias,  $T_A = T_B$ , fixation of cooperation occurs if  $\alpha$  is high enough,  $\frac{c}{b} \cdot \frac{1-T}{T} < \alpha$  (inequality 16; in this case  $a_1 = a_2$ ).

256 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 258 cooperation is facilitated by oblique transmission, whereas if  $T_A < T_B$ , then evolution of cooperation

is facilitated by vertical transmission (Figures 2c and 2d).

260

Next, we examine the roles of vertical and oblique transmission in the evolution of cooperation.

262 Fixation of cooperation is possible only if the vertical transmission rate is high enough,

$$v > \hat{v} = \frac{T_B - T_A}{1 - T_A}. \quad (20)$$

264 This condition is necessary for fixation of cooperation, but it is not sufficient. If horizontal transmission  
266 is biased for cooperation,  $T_A > T_B$ , cooperation can fix with any vertical transmission rate (because  
268  $\hat{v} < 0$ ). 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 2b and 2d).

With only vertical transmission ( $v = 1$ ), from inequality 15, cooperation increases when rare if

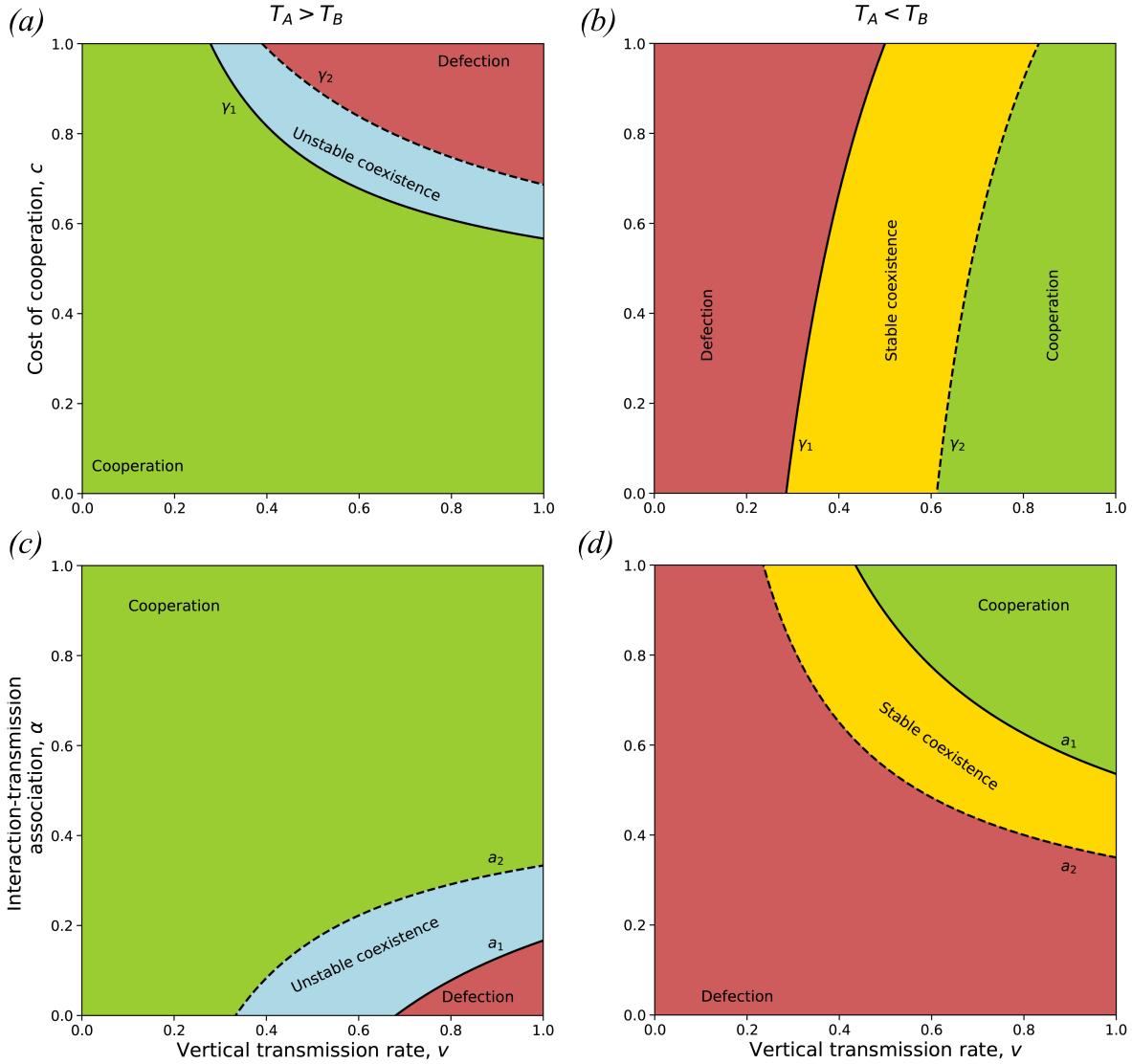
$$270 \quad c < \frac{b\alpha T_A + (T_A - T_B)}{1 - T_B}, \quad (21)$$

which can also be written as

$$272 \quad \frac{c(1 - T_B) - (T_A - T_B)}{bT_A} < \alpha. \quad (22)$$

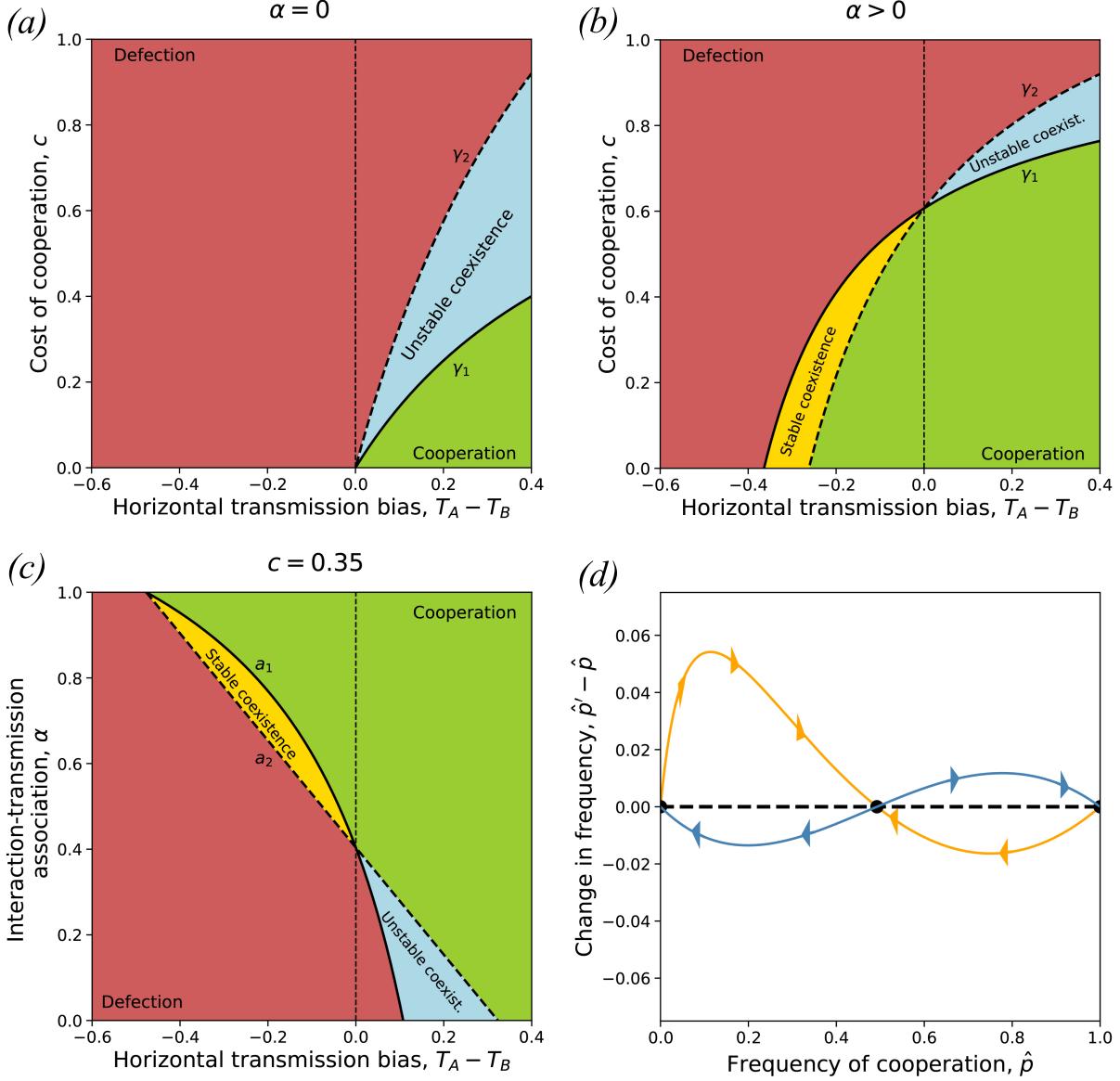
In the absence of vertical transmission ( $v = 0$ ), from recursion 12 we see that the frequency of the  
274 cooperator phenotype among adults increases every generation, i.e.  $p' > p$ , if there is a horizontal  
transmission bias in favor of cooperation, namely  $T_A > T_B$ . That is, if  $v = 0$ , then selection plays no  
276 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,  
278  $T_A = T_B$ , phenotype frequencies do not change,  $\hat{p}' = \hat{p}$ .

Cooperation and defection can coexist at frequencies  $\hat{p}^*$  and  $1 - \hat{p}^*$  (Eq. 13). When it is feasible, this  
280 equilibrium is stable or unstable under the conditions of Result 1, parts 3 and 4, respectively. The  
yellow and blue areas in Figures 3 and 2 show cases of stable and unstable polymorphism, respectively.  
282 When  $\hat{p}^*$  is unstable, cooperation will fix if its initial frequency is  $\hat{p} > \hat{p}^*$ , and defection will fix if  
 $\hat{p} < \hat{p}^*$ .  $\hat{p}^*$  is unstable when there is horizontal transmission bias for cooperation,  $T_A > T_B$ , and the  
284 cost is intermediate,  $\gamma_1 < c < \gamma_2$ . Figure 3d shows  $\hat{p}' - \hat{p}$  as a function of  $\hat{p}$ .



**Figure 2: 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  $\gamma_1$  and  $\gamma_2$  (Eqs. 14) are represented by the solid and dashed lines, respectively. **(c-d)** Interaction-transmission association  $\alpha$  is on the y-axis and the interaction-transmission association thresholds  $a_1$  and  $a_2$  (Eqs. 19) 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 3: 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  $\gamma_1$  and  $\gamma_2$  (Eq. 14) are the solid and dashed lines, respectively. In panel **(c)**, interaction-transmission association  $\alpha$  is on the y-axis and the interaction-transmission association thresholds  $a_1$  and  $a_2$  (Eqs. 19) 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.

## 4.2 Evolution of interaction-transmission association

We now focus on the evolution of interaction-transmission association under perfect vertical transmission,  $v = 1$ , 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. 13). 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. 14), see Figure 3b. 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  $\alpha$  (Appendix C). Therefore,  $\bar{w}^*$  increases as  $\alpha$  increases. But can this population-level advantage lead to the evolution of  $\alpha$ ?

To answer this question, we add a “modifier locus” [31, 32, 33, 34] that determines the value of  $\alpha$  but has no direct effect on fitness. This locus has two alleles,  $M$  and  $m$ , which induce interaction-transmission associations  $\alpha_1$  and  $\alpha_2$ , 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” [33, 35] and obtain the following result.

298

**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  $\alpha$ .*

The analysis is in Appendix D. This reduction principle entails that successful invasions will reduce the frequency of cooperation, as well as the population mean fitness (Figure 4). Furthermore, if we a modifier allele that decreases  $\alpha$  appears and invades the population from time to time, then the value of  $\alpha$  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 3c). Thus, we can expect the value of  $\alpha$  to approach  $a_2$ , the frequency of cooperation to fall to zero, and the population mean fitness to decrease to one (Figure 4).

308 We can do similiar analysis for the general case where  $0 < v < 1$  as we did in Appendix D. First we found the value of the characteristic polynomial at 1,  $R(1)$  using SymPy [36], a Python library for 310 symbolic mathematics.

$$R(1) = \frac{cv\hat{p}^*[T_A b\hat{p}^{*2}(v\alpha_2 - 1) - 2T_A b\hat{p}^*v\alpha_2 + T_A b\hat{p}^*(1 + v\alpha_2)]}{b\hat{p}(b\hat{p}^* - 2\hat{p}^* + 2) + c\hat{p}^*(c\hat{p}^* - 2) + 1} \\ + \frac{cv\hat{p}^*[T_A c\hat{p}^{*2}(1 - v) + T_A c\hat{p}^*(v - 1) - T_A \hat{p}^*(1 - c) + T_A]}{b\hat{p}(b\hat{p}^* - 2\hat{p}^* + 2) + c\hat{p}^*(c\hat{p}^* - 2) + 1} \\ + \frac{cv\hat{p}^*[T_B b\hat{p}^{*2}(1 - v\alpha_2) + T_B b\hat{p}^2(v\alpha_2 - 1) + T_B c\hat{p}^{*2}(v - 1)]}{b\hat{p}(b\hat{p}^* - 2\hat{p}^* + 2) + c\hat{p}^*(c\hat{p}^* - 2) + 1} \\ + \frac{cv\hat{p}^*[T_B c\hat{p}^*(1 - v) + T_B cv(1 - \hat{p}^* + T_B(\hat{p}^* - 1) + cv(\hat{p}^* - 1))]}{b\hat{p}(b\hat{p}^* - 2\hat{p}^* + 2) + c\hat{p}^*(c\hat{p}^* - 2) + 1} \quad (23)$$

312 Similarly to what we did in Appendix D, we should find when  $R(1) < 0$ . However, when adding the vertical transmission into the system the math is more complicated. Therefore, we will use a different 314 technique. We know that when  $\alpha_1 = \alpha_2$ ,  $R(1) = 0$ . Now, let’s assume that  $\alpha_2 = \alpha_1 + \epsilon$ . The derivative

of  $R(1)$  by  $\epsilon$  will give us better understand about the sign of  $R(1)$  when  $\alpha_2 < \alpha_1$ .

$$316 \quad \frac{\partial R(1)}{\partial \epsilon} = \frac{cbv^2 \hat{p}^* [T_A \hat{p}^{*2} - 2T_A \hat{p}^* + T_A - T_B \hat{p}^{*2} + T_B \hat{p}^*]}{b \hat{p}^* (b \hat{p}^* - 2\hat{p}^* + 2) + c \hat{p}^* (c \hat{p}^* - 2) + 1} \quad (24)$$

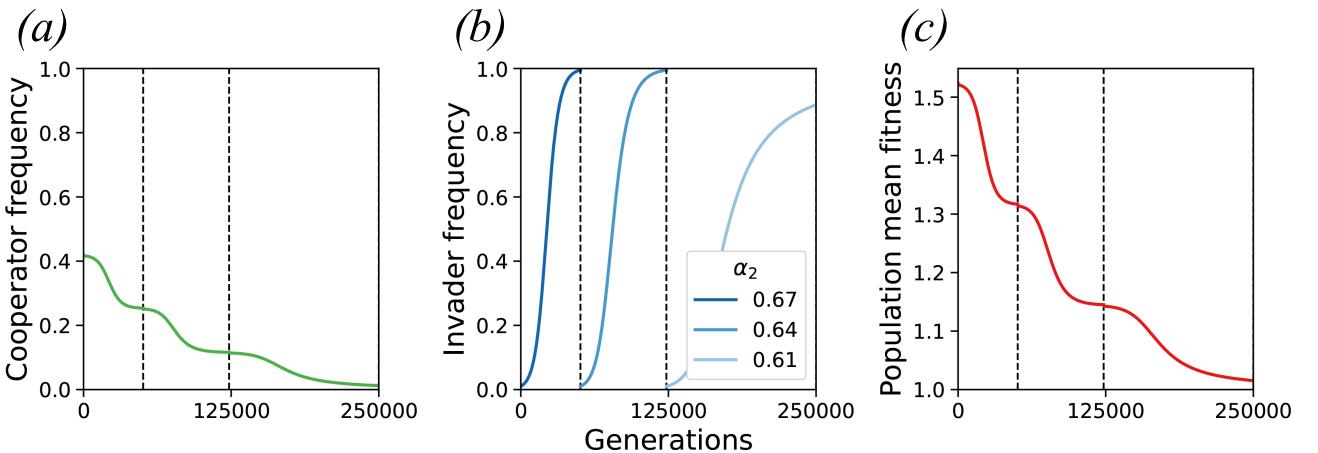
We can simplify Eq. 24

$$318 \quad \frac{\partial R(1)}{\partial \epsilon} = \frac{cbv^2 \hat{p}^* [(T_A - T_B) \hat{p}^{*2} + (T_B - 2T_A) \hat{p}^* + T_A]}{(b \hat{p}^* - c \hat{p}^*)^2 + 2\hat{p}^*(b - c) + 1} \quad (25)$$

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

$$322 \quad \hat{p}_1^* = \frac{T_A}{T_A - T_B} \hat{p}_2^* = 1 \quad (26)$$

$\hat{p}_1^* < 0$  since  $T_B > T_A$  and since the quadratic polynomial has negative leading coefficient then the numerator is positive for every  $\hat{p}_1^* < \hat{p}^* < 1$  which is always true. We found that derivative of  $R(1)$   
324 by  $\epsilon$  is positive for every  $\epsilon$ . Therefore,  $R(1)$  grows as  $\epsilon$  grows, or in other words as  $\alpha_2 - \alpha_1$  grows.  
Thus,  $R(1) < 0$  if and only if  $\alpha_2 < \alpha_1$ . Therefore, different value of vertical transmission does not  
326 change the evolution of social-interaction association and invasion is possible when  $\alpha_2 < \alpha_1$ .



**Figure 4: Reduction principle for interaction-transmission association.** Consecutive fixation of modifier alleles that reduce interaction-transmission association  $\alpha$  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  $\alpha_2 = 0.605$ .

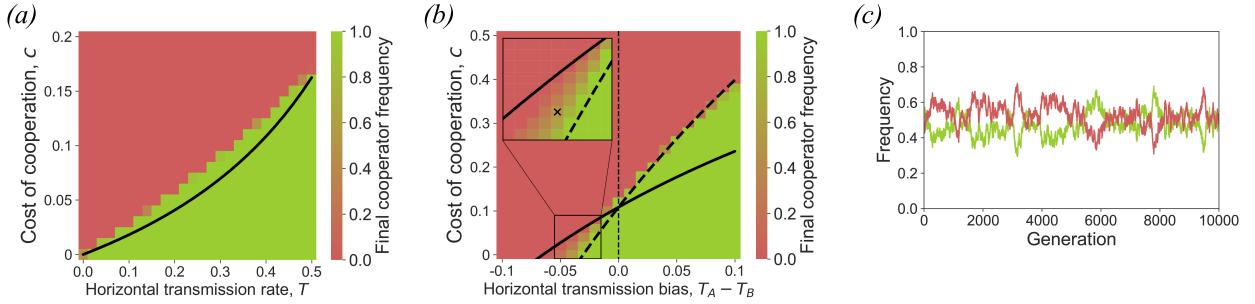
### 4.3 Population structure

328 All the simulations in this section were made by Ohad Lewin-Epstein from Tel Aviv University.  
Interaction-transmission association may also emerge from population structure. Consider a pop-  
330 ulation 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 [24]. Each individual is characterized  
332 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  
334 individuals are randomly chosen to "initiate" interactions, and these initiators interact with a random  
neighbor (i.e. individual in a neighboring site) in a prisoners' dilemma game (Figure 1c) and a random  
336 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  
338 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  $\alpha$  in this model is the probability that  
340 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  
342 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.

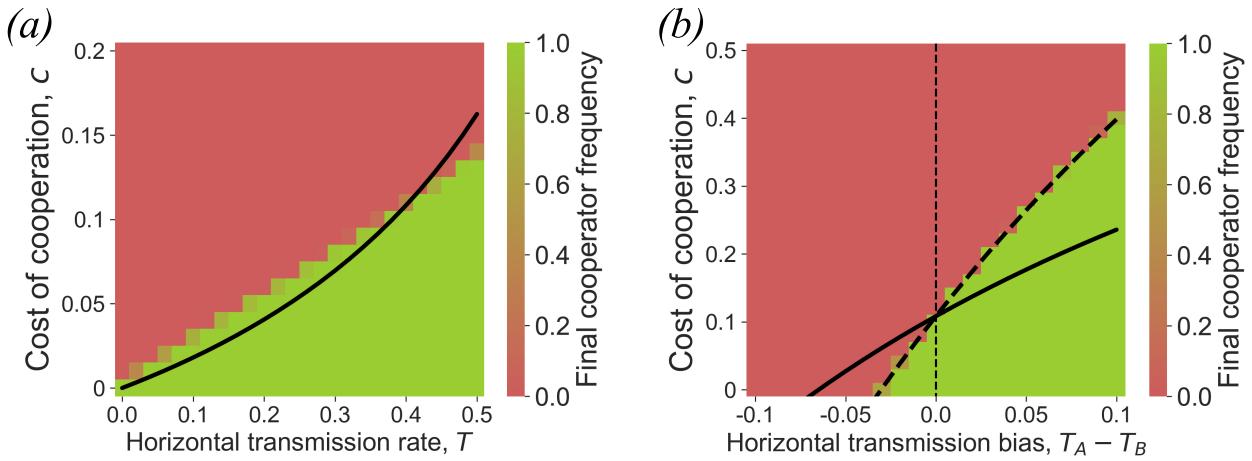
344 The order of the interactions across the grid at each generation is random. After all interactions take  
place, an individual's fitness is determined by  $w = 1 + b \cdot n_b - c \cdot n_c$ , where  $n_b$  is the number of interactions  
346 that individual had with cooperative neighbors, and  $n_c$  is the number of interactions in which that  
individual cooperated (note that the phenotype may change between consecutive interactions due to  
348 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  
350 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  
352 the same phenotype as their parents (i.e.  $v = 1$ ).

The outcomes of stochastic simulations with such a structured population are shown in Figure 5, which  
354 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  
356 example of stochastic stable polymorphism is shown in Figure 5c. Changing the simulation so that  
selection is local (i.e. sites can only be settled by offspring of neighboring parents) had only a minor  
358 effect on the agreement with the derived conditions (Figure 6).

These comparisons between the deterministic unstructured model and the stochastic structured model  
360 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  
362 for interaction-transmission association,  $\alpha$ , can represent local interactions between individuals.



**Figure 5: 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 16 for panel (a) and in Eqs. 14 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 [38]. 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 6: 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,  $0.3 < T_B < 0.5$ .

## 5 Discussion

364 Under a combination of vertical, oblique, and horizontal transmission with payoffs in the form  
365 of a prisoner’s dilemma game, cooperation or defection can either fix or coexist, depending on  
366 the relationship between the cost and benefit of cooperation, the horizontal transmission bias, and  
367 the association between social interaction and horizontal transmission (Result 1, Figures 2 and 3).  
368 Importantly, cooperation can increase when initially rare (i.e. invade a population of defectors) if and  
369 only if, rewriting inequality 15,  $c \cdot v(1 - T_B) < b \cdot v\alpha T_A + (T_A - T_B)$ , namely, the effective cost of  
370 cooperation (left-hand side) is smaller than the effective benefit plus the horizontal transmission bias  
371 (right-hand side). This condition cannot be formulated in the form of Hamilton’s rule,  $c < b \cdot r$ , due to  
372 the effect of biased horizontal transmission, represented by  $(T_A - T_B)$ . Remarkably, a polymorphism  
373 of cooperation and defection can be stable if horizontal transmission is biased in favor of defection  
374 ( $T_A < T_B$ ) and both  $c$  and  $\alpha$  are intermediate (yellow areas in Figures 2 and 3).

We find that stronger interaction-transmission association  $\alpha$  leads to evolution of higher frequency  
376 of cooperation and increased population mean fitness. Nevertheless, when cooperation and defection  
377 coexist,  $\alpha$  is expected to be reduced by natural selection, leading to extinction of cooperation and  
378 decreased population mean fitness (Result 2, Figure 4). With  $\alpha = 0$ , the benefit of cooperation cannot  
facilitate its evolution; it can only succeed if horizontal transmission is biased in its favor.

380 Indeed, in our model, horizontal transmission plays a major role in the evolution of cooperation:  
381 increasing the transmission of cooperation,  $T_A$ , or decreasing the transmission of defection,  $T_B$ , facil-  
382 tates 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  
384 oblique transmission,  $1 - v$ , will facilitate the evolution of cooperation. In contrast, when the bias is  
385 in favor of defection,  $T_A < T_B$ , higher rates of vertical transmission,  $v$ , are advantageous for cooper-  
386 ation, and the rate of vertical transmission must be high enough ( $v > \hat{v}$ ) for cooperation to fix in the  
population.

388 Our deterministic model provides a good approximation to outcomes of simulations of a complex  
389 stochastic model with population structure in which individuals can only interact with and transmit  
390 to their neighbors. In these structured populations interaction-transmission association arises due to  
both social interactions and horizontal cultural transmission being local (Figure 5).

392 Feldman et al. [21] studied the dynamics of an altruistic phenotype with vertical cultural transmission  
393 and a gene that modifies the transmission of the phenotype. Their results are very sensitive to  
394 this genetic modification: without it, the conditions for invasion of the altruistic phenotype reduce  
395 to Hamilton’s rule. Further work is needed to incorporate such genetic modification of cultural  
396 transmission into our model. Woodcock [22] stressed the significance of non-vertical transmission for  
the evolution of cooperation and carried out simulations with prisoner’s dilemma payoffs but without  
398 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  
400 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.

- 402 To understand the role of horizontal transmission, we first review the role of *assortment*. Eshel and  
404 Cavalli-Sforza [18] showed that altruism can evolve when the tendency for *assortative meeting*, i.e.  
406 for individuals to interact with others of their own phenotype, is strong enough. Fletcher and Doebeli  
408 [17] further argued that a general explanation for the evolution of altruism is given by *assortment*: the  
410 correlation between individuals that carry an altruistic trait and the amount of altruistic behavior in  
their interaction group (see also Bijma and Aanen [19]). They suggested that to explain the evolution of  
412 altruism, we should seek mechanisms that generate assortment, such as population structure, repeated  
interactions, and individual recognition. Our results highlight another mechanism for generating  
414 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 ( $T_A < T_B$ ).
- 416 How does non-vertical transmission generate assortment? Lewin-Epstein et al. [23] and Lewin-  
Epstein and Hadany [24] suggested that microbes that induce their hosts to act altruistically can  
418 be favored by selection, which may help to explain the evolution of cooperation. From the kin  
selection point-of-view, if microbes can be transmitted *horizontally* from one host to another during  
420 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.  
422 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  
424 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  
426 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.  
428 Thus, with horizontal transmission, “assortment between focal cooperative players and cooperative  
acts in their interaction environment” [17] is generated not because the helper is likely to be helped,  
430 but rather because the helped is likely to become a helper.

## References

- 432 [1] Robert Axelrod and William D Hamilton. The evolution of cooperation. *Science*, 211(4489):1390–1396, 1981.
- 434 [2] Lee Alan Dugatkin. *Cooperation among Animals: An Evolutionary Perspective*. Oxford University Press on Demand, 1997.
- 436 [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.
- 438 [4] George E Rice and Priscilla Gainer. “Altruism” in the albino rat. *Journal of Comparative and Physiological Psychology*, 55(1):123, 1962.
- 440 [5] Peter B Stacey and Walter D Koenig, editors. *Cooperative breeding in birds: long term studies of ecology and behaviour*. Cambridge University Press, 1990.
- 442 [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.
- 444 [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.
- 446 [8] J. B. S. Haldane. *The Causes of Evolution*. Longmans, London, 1932.
- 450 [9] William D Hamilton. The genetical evolution of social behaviour. ii. *Journal of Theoretical Biology*, 7(1):17–52, 1964.
- 452 [10] Edward O Wilson. Kin selection as the key to altruism: its rise and fall. *Social Research*, pages 159–166, 2005.
- 454 [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.
- 456 [12] Martin A Nowak. Five rules for the evolution of cooperation. *Science*, 314(5805):1560–1563, 2006.
- 458 [13] Luigi Luca Cavalli-Sforza and Marcus W Feldman. *Cultural transmission and evolution: A quantitative approach*. Number 16. Princeton University Press, 1981.
- 460 [14] Peter J Richerson and Robert Boyd. *Not by Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press, 2008.
- 462 [15] Stephen J Lycett and John AJ Gowlett. On questions surrounding the acheulean ‘tradition’. *World Archaeology*, 40(3):295–315, 2008.

- 464 [16] 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–  
466 E1183, 2018.
- 468 [17] 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.
- 470 [18] Ilan Eshel and Luigi Luca Cavalli-Sforza. Assortment of encounters and evolution of coopera-  
tiveness. *Proceedings of the National Academy of Sciences*, 79(4):1331–1335, 1982.
- 472 [19] Piter Bijma and Duur K. Aanen. Assortment, Hamilton’s rule and multilevel selection. *Proc. R.  
Soc. B Biol. Sci.*, 277(1682):673–675, 2010.
- 474 [20] 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.
- 476 [21] 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.
- 478 [22] Scott Woodcock. The significance of non-vertical transmission of phenotype for the evolution  
of altruism. *Biology and Philosophy*, 21(2):213–234, 2006.
- 480 [23] Ohad Lewin-Epstein, Ranit Aharonov, and Lilach Hadany. Microbes can help explain the  
evolution of host altruism. *Nature Communications*, 8:14040, 2017.
- 482 [24] 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.
- 484 [25] 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.
- 486 [26] Matthew R. Zefferman. Mothers teach daughters because daughters teach granddaughters: the  
evolution of sex-biased transmission. *Behav. Ecol.*, 27(4):1172–1181, 2016.
- 488 [27] Manfred Milinski. Tit for tat in sticklebacks and the evolution of cooperation. *nature*, 325(6103):  
433–435, 1987.
- 490 [28] Martin A Nowak and Karl Sigmund. Tit for tat in heterogeneous populations. *Nature*, 355(6357):  
250–253, 1992.
- 492 [29] Brian Skyrms. *The stag hunt and the evolution of social structure*. Cambridge University Press,  
2004.
- 494 [30] 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  
496 in the evolution of human cooperation: The interdependence hypothesis. *Current anthropology*,  
53(6):000–000, 2012.

- 498 [31] Marcus W. Feldman. Selection for linkage modification: I. Random mating populations. *Theor. Popul. Biol.*, 3:324–346, 1972.
- 500 [32] Uri Liberman and Marcus W. Feldman. A general reduction principle for genetic modifiers of recombination. *Theor. Popul. Biol.*, 30(3):341–71, dec 1986.
- 502 [33] Uri Liberman and Marcus W. Feldman. Modifiers of mutation rate: A general reduction principle. *Theor. Popul. Biol.*, 30:125–142, 1986.
- 504 [34] Uri Liberman. External stability and ESS: criteria for initial increase of new mutant allele. *J. Math. Biol.*, 26(4):477–485, 1988.
- 506 [35] 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):  
508 E2392–E2400, mar 2017.
- [36] Aaron Meurer, Christopher P Smith, Mateusz Paprocki, Ondřej Čertík, Sergey B Kirpichev,  
510 Matthew Rocklin, AMiT Kumar, Sergiu Ivanov, Jason K Moore, Sartaj Singh, et al. Sympy:  
symbolic computing in python. *PeerJ Computer Science*, 3:e103, 2017.
- 512 [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.
- 514 [38] 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.

# 516 Appendices

## Appendix A Local stability criterion

518 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

$$520 \quad p' < p \Leftrightarrow f(p)/p < 0 \Leftrightarrow \frac{f'(0) \cdot p + O(p^2)}{p} < 0 \Leftrightarrow f'(0) + O(p) < 0. \quad (\text{A1})$$

Therefore, by definition of big-O notation, if  $f'(0) < 0$  then there exists  $\epsilon > 0$  such that for any local  
522 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

$$524 \quad 1 - p' < 1 - p \Leftrightarrow -\frac{f(p)}{1 - p} < 0 \Leftrightarrow \frac{f'(1)(p - 1) + O((p - 1)^2)}{p - 1} < 0 \Leftrightarrow f'(1) - O(1 - p) < 0. \quad (\text{A2})$$

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

## Appendix B Equilibria and stability

528 Let  $f(\hat{p}) = \bar{w}(\hat{p}' - \hat{p})$ . Then, using SymPy [36], a Python library for symbolic mathematics, this simplifies to

$$530 \quad f(\hat{p}) = \bar{w}(\hat{p}' - \hat{p}) = \beta_1 \hat{p}^3 + \beta_2 \hat{p}^2 + \beta_3 \hat{p} , \quad (\text{B1})$$

where

$$\begin{aligned} 532 \quad \beta_1 &= [c(1 - v) - b(1 - \alpha v)](T_A - T_B) , \\ \beta_2 &= -\beta_1 - \beta_3 , \\ \beta_3 &= \alpha b v T_A - c v (1 - T_B) + (T_A - T_B) . \end{aligned} \quad (\text{B2})$$

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

$$534 \quad f(\hat{p}) = \hat{p}(1 - \hat{p})[\alpha b v T - c v (1 - T)] . \quad (\text{B3})$$

536 Clearly the only two equilibria are the fixations  $\hat{p} = 0$  and  $\hat{p} = 1$ , which are locally stable if  $f'(\hat{p}) < 0$  near the equilibrium (see Appendix A), where  $f'(\hat{p}) = (1 - 2\hat{p})[\alpha b v T - c v (1 - T)]$ , so 538 that

$$\begin{aligned} f'(0) &= \alpha b v T - c v (1 - T) , \\ f'(1) &= -\alpha b v T + c v (1 - T) . \end{aligned} \quad (\text{B4})$$

540 In the general case where  $T_A \neq T_B$ , the coefficient  $\beta_1$  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 542 is

$$\hat{p}^* = \frac{\beta_3}{\beta_1} = \frac{\alpha b v T_A - c v (1 - T_B) + (T_A - T_B)}{[c(1 - v) - b(1 - \alpha v)](T_A - T_B)} . \quad (\text{B5})$$

544 Note that the sign of the cubic (Eq. B1) at positive (negative) infinity is equal (opposite) to the sign of  $\beta_1$ . If  $T_A > T_B$ , then

$$546 \quad \beta_1 < [c(1 - \alpha v) - b(1 - \alpha v)](T_A - T_B) = (1 - \alpha v)(c - b)(T_A - T_B) < 0 , \quad (\text{B6})$$

since  $c < b$  and  $\alpha v < 1$ . Hence the signs of the cubic at positive and negative infinity are negative and 548 positive, respectively. First, if  $\beta_3 < \beta_1$  then  $1 < \hat{p}^*$ . Also,  $f'(0) < 0$  and  $f'(1) > 0$ ; that is, fixation 550 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 552 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.

554 Similarly, if  $T_A < T_B$ , then

$$\beta_1 > [c(1 - \alpha v) - b(1 - \alpha v)](T_A - T_B) = (1 - \alpha v)(c - b)(T_A - T_B) > 0 , \quad (\text{B7})$$

556 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

558 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  
560 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  
562 equilibrium.

This analysis can be summarized as follows:

- 564 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  $\beta_1 < \beta_3$ .
- 566 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  $\beta_3 < 0$ .
- 568 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  $\beta_1 < \beta_3 < 0$ .

570 We now proceed to use the cost thresholds,  $\gamma_1$  and  $\gamma_2$ , and the vertical transmission threshold,  $\hat{v}$  (Eq. 14).  
First, assume  $T_A < T_B$ .  $\beta_3 < 0$  requires  $\gamma_1 < c$ . For  $\beta_3 < \beta_1$  we need  $c[v(1-T_B)+(1-v)(T_A-T_B)] >$   
572  $b\alpha T_B + (1+b)(T_A - T_B)$ . 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  
574  $v > \hat{v}$  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  $c < 1 < \gamma_2$ .

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

This analysis is summarized in Result 1.

580 **Appendix C Effect of interaction-transmission association on mean fitness**

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

$$584 \quad \frac{\partial \hat{p}^*}{\partial \alpha} = \frac{bT_A - c(1 - T_B) + (T_A - T_B)}{b(1 - \alpha)^2(T_B - T_A)}. \quad (\text{C1})$$

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

$$586 \quad c < \gamma_1 = \frac{b\alpha T_A + (T_A - T_B)}{1 - T_B} < \frac{bT_A + (T_A - T_B)}{1 - T_B}. \quad (\text{C2})$$

Therefore, the numerator in Eq. C1 is positive. Since  $T_A < T_B$ , the denominator in Eq. C1 is also  
588 positive, and hence the derivative  $\partial \hat{p}^*/\partial \alpha$  is positive. Thus, the population mean fitness increases as interaction-transmission association  $\alpha$  increases.

## 590 Appendix D Reduction principle

We assume here that  $v = 1$ , i.e. no oblique transmission, and therefore  $\hat{p} = \dot{p}$ . Denote the frequencies 592 of the pheno-genotypes  $AM$ ,  $BM$ ,  $Am$ , and  $Bm$  by  $\mathbf{p} = (\dot{p}_1, \dot{p}_2, \dot{p}_3, \dot{p}_4)$ . The frequencies of the pheno-genotypes in the next generation are defined by the recursion system,

$$\begin{aligned}
\bar{w}\dot{p}'_1 &= \dot{p}_1x(1+b-c)(1-(1-\alpha_1)(1-x)T_B) + \\
&\quad \dot{p}_1(1-x)(1-c)(1-\alpha_1T_Bx-T_B(1-x)) + \\
&\quad \dot{p}_2x(1+b)T_A(x+\alpha_1(1-x)) + \\
&\quad \dot{p}_2(1-x)x(1-\alpha_1)T_A, \\
\bar{w}\dot{p}'_2 &= \dot{p}_1x(1+b-c)(1-\alpha_1)(1-x)T_B + \\
&\quad \dot{p}_1(1-x)(1-c)(\alpha_1T_B+(1-\alpha_1)(1-x)T_B) + \\
&\quad \dot{p}_2x(1+b)(1-\alpha_1T_A(1-x)-T_AX) + \\
&\quad \dot{p}_2(1-x)(1-(1-\alpha_1)xT_A), \\
\bar{w}\dot{p}'_3 &= \dot{p}_3x(1+b-c)(1-(1-\alpha_2)(1-x)T_B) + \\
&\quad \dot{p}_3(1-x)(1-c)(1-\alpha_2T_Bx-T_B(1-x)) + \\
&\quad \dot{p}_4x(1+b)T_A(x+\alpha_2(1-x)) + \\
&\quad \dot{p}_4(1-x)x(1-\alpha_2)T_A, \\
\bar{w}\dot{p}'_4 &= \dot{p}_3x(1+b-c)(1-\alpha_2)(1-x)T_B + \\
&\quad \dot{p}_3(1-x)(1-c)(\alpha_2T_B+(1-\alpha_2)(1-x)T_B) + \\
&\quad \dot{p}_4x(1+b)(1-\alpha_2T_A(1-x)-T_AX) + \\
&\quad \dot{p}_4(1-x)(1-(1-\alpha_2)xT_A),
\end{aligned} \tag{D1}$$

594

where  $x = \dot{p}_1 + \dot{p}_3$  is the total frequency of the cooperative phenotype  $A$ , and  $\bar{w} = 1 + (b - c)x$  is the 596 population mean fitness.

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

$$598 \quad \dot{p}^* = \frac{c(1 - T_B) - b\alpha_1 T_A - (T_A - T_B)}{b(1 - \alpha_1)(T_A - T_B)}, \tag{D2}$$

setting  $\alpha = \alpha_1$  and  $v = 1$  in Eq. 13. When  $v = 1$ ,  $\dot{p}^*$  is a feasible polymorphism ( $0 < \dot{p}^* < 1$ ) if 600  $T_A < T_B$  and  $\gamma_2 < c < \gamma_1$  (Result 1).

The local stability of  $\mathbf{p}^*$  to the introduction of allele  $m$  is determined by the linear approximation  $\mathbf{L}^*$  602 of the transformation in Eq. D1 near  $\mathbf{p}^*$  (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  $\mathbf{L}_{ex}^*$  604 [33, 35]. The latter is the external stability matrix: the linear approximation to the transformation near  $\mathbf{p}^*$  involving only the pheno-genotypes  $Am$  and  $Bm$ , derived from Eq. D1, with  $\bar{w}^* = 1 + (b - c)\dot{p}^*$  as

606 the stable population mean fitness,

$$\mathbf{L}_{ex}^* = \frac{1}{\bar{w}^*} \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix} = \frac{1}{\bar{w}^*} \begin{bmatrix} \frac{\partial \bar{w} \dot{p}_3'}{\partial \dot{p}_3}(\mathbf{p}^*) & \frac{\partial \bar{w} \dot{p}_3'}{\partial \dot{p}_4}(\mathbf{p}^*) \\ \frac{\partial \bar{w} \dot{p}_4'}{\partial \dot{p}_3}(\mathbf{p}^*) & \frac{\partial \bar{w} \dot{p}_4'}{\partial \dot{p}_4}(\mathbf{p}^*) \end{bmatrix} = \frac{1}{\bar{w}^*} \begin{bmatrix} (1 + b\dot{p}^* - c)(1 - T_B(1 - \dot{p}^*)) + b\dot{p}^*\alpha_2 T_B(1 - \dot{p}^*) & (1 + b\dot{p}^*)T_A \dot{p}^* + b\dot{p}^*\alpha_2 T_A(1 - \dot{p}^*) \\ (1 + b\dot{p}^* - c)T_B(1 - \dot{p}^*) - b\dot{p}^*\alpha_2 T_B(1 - \dot{p}^*) & (1 + b\dot{p}^*)(1 - T_A \dot{p}^*) - b\dot{p}^*\alpha_2 T_A(1 - \dot{p}^*) \end{bmatrix}. \quad (\text{D3})$$

608 Because we assume that  $\mathbf{p}^*$  is internally stable (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  
610 stability matrix  $\mathbf{L}_{ex}^*$ . This is a positive matrix, and due to the Perron-Frobenius theorem, the leading  
612 eigenvalue of  $\mathbf{L}_{ex}^*$  is real and positive. Thus, if the leading eigenvalue is less (greater) than one, then the  
614 equilibrium  $\mathbf{p}^*$  is externally stable (unstable) and allele  $m$  cannot (can) invade the population of allele  
616  $M$ . 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,  $\lim_{\lambda \rightarrow \pm\infty} R(\lambda) = \infty$ , 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. Since  
618 multiplication by a positive factor doesn't change the sign, and using the properties of the determinant,  
we have

$$620 \quad \begin{aligned} \text{sign } R(1) &= \text{sign } \det(\mathbf{L}_{ex}^* - \mathbf{I}) = \text{sign } (\bar{w}^*)^2 \det(\mathbf{L}_{ex}^* - \mathbf{I}) = \\ \text{sign } \det(\bar{w}^* \mathbf{L}_{ex}^* - \bar{w}^* \mathbf{I}) &= \text{sign } \det \begin{bmatrix} l_{11} - \bar{w}^* & l_{12} \\ l_{21} & l_{22} - \bar{w}^* \end{bmatrix}, \end{aligned} \quad (\text{D4})$$

where  $l_{ij}$  are defined in Eq. D3. Adding the second row in Eq. D4 to the first row, which does not  
622 change the determinant, and substituting  $\bar{w}^* = 1 + (b - c)\dot{p}^*$ , we get

$$\begin{aligned} \text{sign } R(1) &= \text{sign } \det \begin{bmatrix} -c(1 - \dot{p}^*) & c\dot{p}^* \\ (1 - \dot{p}^*)[(1 + b\dot{p}^* - c)T_B - b\alpha_2 T_B \dot{p}^*] & \dot{p}^*[-(1 + b\dot{p}^*)T_A - b\alpha_2 T_A(1 - \dot{p}^*) + c] \end{bmatrix} = \\ \text{sign } &\left[ c\dot{p}^*(1 - \dot{p}^*) \cdot \det \begin{bmatrix} -1 & 1 \\ (1 + b\dot{p}^* - c)T_B - b\alpha_2 T_B \dot{p}^* & -(1 + b\dot{p}^*)T_A - b\alpha_2 T_A(1 - \dot{p}^*) + c \end{bmatrix} \right] = \\ \text{sign } \det &\begin{bmatrix} -1 & 1 \\ (1 + b\dot{p}^* - c)T_B - b\alpha_2 T_B \dot{p}^* & -(1 + b\dot{p}^*)T_A - b\alpha_2 T_A(1 - \dot{p}^*) + c \end{bmatrix}, \end{aligned} \quad (\text{D5})$$

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

$$\begin{aligned} \text{sign } R(1) &= \text{sign } \left[ (1 + b\dot{p}^*)T_A + b\alpha_2 T_A(1 - \dot{p}^*) - c - (1 + b\dot{p}^* - c)T_B + b\dot{p}^*\alpha_2 T_B \right] = \\ \text{sign } &\left[ (1 + b(1 - \alpha_2)\dot{p}^*)(T_A - T_B) + b\alpha_2 T_A - c(1 - T_B) \right]. \end{aligned} \quad (\text{D6})$$

626 Substituting  $\dot{p}^*$  from Eq. D2, we get

$$\begin{aligned}
R(1) < 0 \Leftrightarrow & [c(1 - T_B) - b\alpha_1 T_A - (T_A - T_B)] \frac{1 - \alpha_2}{1 - \alpha_1} - c(1 - T_B) + b\alpha_2 T_A + (T_A - T_B) < 0 \Leftrightarrow \\
& (1 - \alpha_2)[c(1 - T_B) - b\alpha_1 T_A - (T_A - T_B)] < (1 - \alpha_1)[c(1 - T_B) - b\alpha_2 T_A - (T_A - T_B)] \Leftrightarrow \\
& -b\alpha_1 T_A - \alpha_2 c(1 - T_B) + \alpha_2 (T_A - T_B) < -b\alpha_2 T_A - \alpha_1 c(1 - T_B) + \alpha_1 (T_A - T_B) \Leftrightarrow \\
& \alpha_1[c(1 - T_B) - bT_A - (T_A - T_B)] < \alpha_2[c(1 - T_B) - bT_A - (T_A - T_B)] \Leftrightarrow \\
& \alpha_1[bT_A + (T_A - T_B) - c(1 - T_B)] > \alpha_2[bT_A + (T_A - T_B) - c(1 - T_B)] . 
\end{aligned} \tag{D7}$$

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

$$\begin{aligned}
c < \gamma_1 = \frac{b\alpha_1 T_A + (T_A - T_B)}{1 - T_B} \Leftrightarrow \\
0 < b\alpha_1 T_A + (T_A - T_B) - c(1 - T_B) \Rightarrow \\
0 < bT_A + (T_A - T_B) - c(1 - T_B) .
\end{aligned} \tag{D8}$$

630 Combining inequalities D7 and D8, we find that  $R(1) < 0$  if and only if  $\alpha_1 > \alpha_2$ , which is a sufficient condition for external instability. Therefore, if  $\alpha_2$ , the interaction-transmission association of the 632 invading modifier allele  $m$ , is less than  $\alpha_1$ , the interaction-transmission association of the resident allele  $M$ , then invasion will be successful.

634 Determining a necessary and sufficient condition for successful invasion is more complicated, requiring analysis of the sign of  $R'(1)$ . However, we have numerically validated that the leading eigenvalue is 636 greater than one if and only if  $\alpha_1 > \alpha_2$ .