

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

Dor Cohen¹, Ohad Lewin-Epstein², Marcus W. Feldman³, and Yoav Ram^{1,4,5,*}

¹School of Computer Science, Interdisciplinary Center Herzliya, Herzliya, Israel

²School of Plant Sciences and Food Security, Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel

³Department of Biology, Stanford University, Stanford, CA

⁴School of Zoology, Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel

⁵Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

*Corresponding author: yoav@yoavram.com

July 11, 2021

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.

Introduction

22 Cooperative behavior can reduce an individual's fitness and increase the fitness of its conspecifics or
24 competitors [1]. Nevertheless, cooperative behavior appears to occur in many animals [2], including
26 humans, primates [3], rats [4], birds [5, 6], and lizards [7]. Evolution of cooperative behavior has
28 been an important focus of research in evolutionary theory since at least the 1930s [8]. Since the work
30 of Hamilton [9] and Axelrod and Hamilton [1], theories for the evolution of cooperative and altruistic
32 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 benefit to the recipient, b , times the
relatedness, r , between the recipient and the actor. This condition is known as *Hamilton's rule*:

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

34 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.

36 Eshel and Cavalli-Sforza [10] 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
38 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
40 population structure or active partner choice. In their model, cooperative behavior can evolve if [10,
eq. 3.2]

$$c < b \cdot m, \quad (2)$$

42 where b and c are the benefit and cost of cooperation¹.

44 The role of assortment in the evolution of altruism was emphasized by Fletcher and Doebeli [11].
They found that in a *public-goods* game, altruism will evolve if cooperative individuals experience
46 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
48 environment." With some change in parameters, this condition is summarized by [11, eq. 2.3]

$$c < b \cdot (p_C - p_D), \quad (3)$$

50 where p_C is the probability that a cooperator receives help, and p_D is the probability that a defector
receives help². Bijma and Aanen [12] obtained a result related to inequality 3 for other games.

52 Cooperation can also evolve when interactions are determined by population structure. For example,
Ohtsuki et al. [13] studied populations on graphs with average degree k , that is, the average individual
54 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

$$c < b \cdot \frac{1}{k}. \quad (4)$$

56 They thus interpret $1/k$ as *social relatedness* or *social viscosity* [13].

58 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,

¹In 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 [10, eq. 4.6].

²Inequality 3 generalizes inequalities 1 and 2 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.

60 or other modes of communication [14, 15]. Feldman et al. [16] introduced the first model for the
 62 evolution of altruism by cultural transmission with kin selection and demonstrated that if the fidelity
 64 of cultural transmission of altruism is φ , then the condition for evolution of altruism in the case of
 sib-to-sib altruism is [16, Eq. 16]

$$64 \quad c < b \cdot \varphi - \frac{1 - \varphi}{\varphi}. \quad (5)$$

In inequality 5, φ replaces relatedness (r in inequality 1) or assortment (m in inequality 2), 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).

68 Cultural transmission may be modeled as vertical, horizontal, or oblique: vertical transmission occurs
 70 between parents and offspring, horizontal transmission occurs between individuals from the same
 72 generation, and oblique transmission occurs to offspring from the generation to which their parents
 74 belong (i.e. from non-parental adults). Evolution under either of these transmission models can be
 76 more rapid than under pure vertical transmission [14, 17, 18]. Both Woodcock [19] and Lewin-Epstein
 et al. [20] 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 [20, 21, 22] are mathematically similar to models of cultural transmission, and they
 also emphasize the role of non-vertical transmission [14].

78 Here, we study models for the cultural evolution of cooperation that include both vertical and non-
 80 vertical transmission. In our models behavioral changes are mediated by cultural transmission that
 82 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
 84 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
 86 with social interactions, can enhance the evolution of cooperation even when genetic transmission
 cannot, partly because it facilitates the generation of assortment [11], and partly because it diminishes
 the effect of selection (due to non-vertical transmission from non-reproducing individuals [18]).

Models

88 Consider a very large well-mixed 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 cultural
 90 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. [18], given that the
 92 parent's phenotype is ϕ and assuming uni-parental inheritance [23], the conditional probability that
 the phenotype ϕ' of the offspring is A is

$$94 \quad P(\phi' = A | \phi) = \begin{cases} v + (1 - v)p, & \text{if } \phi = A \\ (1 - v)p, & \text{if } \phi = B \end{cases}, \quad (6)$$

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

96 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
 98 transmission) is

$$\hat{p} = \dot{p}[v + (1 - v)p] + (1 - \dot{p})[(1 - v)p] = v\dot{p} + (1 - v)p. \quad (7)$$

100 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

102 b , where we assume $c < b$ (i.e. donation game). Figure 1c shows the payoff matrix: the fitness of an individual with phenotype ϕ_1 when interacting with a partner of phenotype ϕ_2 .

104 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
106 different phenotypes interact with probability $2\hat{p}(1 - \hat{p})$. Horizontal cultural transmission occurs
108 between pairs of individuals from the same generation. It occurs between socially interacting partners
110 with probability α , or between a random pair with probability $1 - \alpha$ (see Figure 1b). However,
112 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

$$\begin{aligned} p' &= \hat{p}^2 [\alpha + (1 - \alpha)(\hat{p} + (1 - \hat{p})(1 - T_B))] + \\ &\quad \hat{p}(1 - \hat{p}) [\alpha(1 - T_B) + (1 - \alpha)(\hat{p} + (1 - \hat{p})(1 - T_B))] + \\ &\quad (1 - \hat{p})\hat{p} [\alpha T_A + (1 - \alpha)\hat{p}T_A] + (1 - \hat{p})^2 [(1 - \alpha)\hat{p}T_A] \\ &= \hat{p}^2(T_B - T_A) + \hat{p}(1 + T_A - T_B). \end{aligned} \quad (8)$$

114 For example, the first term in Eq. 8 describes the case where two juveniles with phenotype A interact
with probability \hat{p}^2 . In this case, the focal individual will retain its phenotype if (i) its social interaction
116 partner is also its horizontal transmission partner, with probability α ; or (ii) its horizontal transmission
partner is another individual, with probability $(1 - \alpha)$, and (ii.a) that individual also has phenotype A ,
118 with probability \hat{p} , or (ii.b) that individual has phenotype B , with probability $(1 - \hat{p})$, but horizontal
transmission is unsuccessful, with probability $(1 - T_B)$. The frequency of A among parents follows a
120 similar dynamic but must also include the effect of natural selection. Therefore, each right-hand term
from Eq. 8 is multiplied by the corresponding fitness value (Table 1, Figure 1c), which depends on the
122 phenotypes of the two interaction partners. Therefore, the frequency of phenotype A among parents
is

$$\begin{aligned} \bar{w}p' &= \hat{p}^2(1 + b - c)[\alpha + (1 - \alpha)(\hat{p} + (1 - \hat{p})(1 - T_B))] + \\ 124 &\quad \hat{p}(1 - \hat{p})(1 - c)[\alpha(1 - T_B) + (1 - \alpha)(\hat{p} + (1 - \hat{p})(1 - T_B))] + \\ &\quad (1 - \hat{p})\hat{p}(1 + b)[\alpha T_A + (1 - \alpha)\hat{p}T_A] + (1 - \hat{p})^2[(1 - \alpha)\hat{p}T_A], \end{aligned} \quad (9)$$

where fitness values are taken from Figure 1c and Table 1, and the population mean fitness is
126 $\bar{w} = 1 + \hat{p}(b - c)$. Starting from Eq. 7 with $\hat{p}' = v\hat{p}' + (1 - v)p'$, we substitute p' from Eq. 8 and p'
from Eq. 9 and obtain

$$\begin{aligned} 128 \quad \hat{p}' &= \frac{v}{\bar{w}} \left[\hat{p}^2(1 + b - c)(1 - (1 - \hat{p})(1 - \alpha)T_B) \right] + \\ &\quad \frac{v}{\bar{w}} \left[\hat{p}(1 - \hat{p})(1 - c)(\hat{p}(1 - \alpha)T_B + 1 - T_B) \right] + \\ &\quad \frac{v}{\bar{w}} \left[\hat{p}(1 - \hat{p})(1 + b)(\hat{p}(1 - \alpha) + \alpha)T_A \right] + \\ &\quad \frac{v}{\bar{w}} (1 - \hat{p})^2 \hat{p}(1 - \alpha)T_A + (1 - v)\hat{p}^2(T_B - T_A) + (1 - v)\hat{p}(1 + T_A - T_B). \end{aligned} \quad (10)$$

Table 2 lists the model variables and parameters.

130 Results

We determine the equilibria of the model in Eq. 10 and analyze their local stability. We then analyze
132 the evolution of a modifier of interaction-transmission association, α . Finally, we compare derived
conditions to outcomes of stochastic simulations with a structured population.

134 **Evolution of cooperation**

The fixed points (equilibria) of the recursion (Eq. 10) are $\hat{p} = 0$, $\hat{p} = 1$, and (see Eq. B5)

$$136 \quad \hat{p}^* = \frac{\alpha bvT_A - cv(1 - T_B) + (T_A - T_B)}{[c(1 - v) - b(1 - \alpha v)](T_A - T_B)}. \quad (11)$$

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

$$138 \quad \gamma_1 = \frac{bvaT_A + (T_A - T_B)}{v(1 - T_B)}, \quad \gamma_2 = \frac{bvaT_B + (1 + b)(T_A - T_B)}{v(1 - T_B) + (1 - v)(T_A - T_B)}, \quad \hat{v} = \frac{T_B - T_A}{1 - T_A}. \quad (12)$$

Then we have the following result.

140

Result 1. *With vertical, horizontal, and oblique transmission, the cultural evolution of 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. 12) :*

- 144 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 $c < \gamma_2$.
- 146 2. Fixation of defection: if (iii) $T_A \geq T_B$ and $\gamma_2 < c$; or if (iv) $T_A < T_B$ and $\gamma_1 < c$.
- 148 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 $\gamma_2 < c < \gamma_1$.
- 148 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 2a, 2b, 3a, and 3b, and the analysis is in Appendix B. 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.

160 **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,*

$$162 \quad c < \gamma_1 = \frac{bvaT_A + (T_A - T_B)}{v(1 - T_B)}. \quad (13)$$

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

Notably, 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 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 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 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, making it harder for cooperation to increase when rare (Figure 2a).

Importantly, this condition cannot be formulated in the commonly used 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 13, cooperation will take over the population from any initial frequency if the cost is low enough,

$$c < b \cdot \frac{\alpha T}{1 - T}, \quad (14)$$

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 3, where $\alpha T/(1-T)$ is a measure of *cultural relatedness* or *cultural assortment*, respectively, similar to the term *social relatedness* used by Ohtsuki et al. [13]. Note that the right-hand side of inequality 14 equals γ_1 when $T = T_A = T_B$.

From inequality 13, 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

$$c < \frac{T_A - T_B}{v(1 - T_B)}. \quad (15)$$

Figure 3a illustrates this condition (for $v = 1$), which is obtained by setting $\alpha = 0$ in inequality 13. 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 13 shows that with perfect interaction-transmission association ($\alpha = 1$), cooperation will increase when rare if

$$c < \frac{bvT_A + (T_A - T_B)}{v(1 - T_B)}. \quad (16)$$

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 16 applies (Figure 3). This is similar to case of microbe-induced cooperation studied by Lewin-Epstein et al. [20]; 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

$$a_1 = \frac{c \cdot v(1 - T_A) - (T_A - T_B)(1 + b - c)}{b \cdot v \cdot T_B}, \quad a_2 = \frac{c \cdot v(1 - T_B) - (T_A - T_B)}{b \cdot v \cdot T_A}. \quad (17)$$

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

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 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, 210 $\frac{c}{b} \cdot \frac{1-T}{T} < \alpha$ (inequality 14; in this case $a_1 = a_2$).

Interestingly, because the sign of $\partial a_2 / \partial v$ is equal to the sign of $T_A - T_B$, the effect of the vertical 212 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 214 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).

216 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,

$$218 \quad v > \hat{v} = \frac{T_B - T_A}{1 - T_A}. \quad (18)$$

This condition is necessary for fixation of cooperation, but it is not sufficient. If horizontal transmission 220 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 222 only if the vertical transmission rate is high enough: in this case oblique transmission can prevent fixation of cooperation (see Figures 2b and 2d).

224 With only vertical transmission ($v = 1$), from inequality 13, cooperation increases when rare if

$$c < \frac{baT_A + (T_A - T_B)}{1 - T_B}, \quad (19)$$

226 which can also be written as

$$\frac{c(1 - T_B) - (T_A - T_B)}{bT_A} < \alpha. \quad (20)$$

228 In the absence of vertical transmission ($v = 0$), from recursion 10 we see that the frequency of the cooperator phenotype among adults increases every generation, i.e. $p' > p$, if there is a horizontal 230 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 232 by differential horizontal transmission of the two phenotypes. With no bias in horizontal transmission, $T_A = T_B$, phenotype frequencies do not change, $\hat{p}' = \hat{p}$.

234 Cooperation and defection can coexist at frequencies \hat{p}^* and $1 - \hat{p}^*$ (Eq. 11). When it is feasible, this equilibrium is stable or unstable under the conditions of Result 1, parts 3 and 4, respectively. The 236 yellow and blue areas in Figures 3 and 2 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 238 $\hat{p} < \hat{p}^*$. \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 3d shows $\hat{p}' - \hat{p}$ as a function of \hat{p} .

240 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. 11). Note that 242 for a stable polymorphism, there must be horizontal bias for defection, $T_A < T_B$, and an intermediate 244 cost of cooperation, $\gamma_2 < c < \gamma_1$ (Eq. 12), see Figure 3b. The equilibrium population mean fitness is 246 $\bar{w}^* = 1 + \hat{p}^*(b - c)$, which is increasing in \hat{p}^* , and \hat{p}^* is increasing in α (Appendix C). Therefore, \bar{w}^* increases as α increases. But can this population-level advantage lead to the evolution of α ?

248 To answer this question, we add a *modifier locus* [24, 25, 26, 27] that determines the value of α but has no direct effect on fitness. This locus has two alleles, M and m , which induce interaction- 250 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* [26, 28]) and obtain the following result.

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 .*

The analysis is in Appendix D. This *reduction principle* [24, 28] entails that successful invasions will reduce the frequency of cooperation, as well as the population mean fitness (Figure S1). Furthermore, if 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 3c). 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 S1). Note that α controls how often an individual learns from its interaction partner. However, from the *phenotype-centred view*, there is no incentive to do so: a cooperator interacting with a defector will not only pay the cost of cooperation but will also risk being "converted" to defection (with probability T_B), whereas a defector interacting with a cooperator will forfeit (with probability T_A) the benefit it received.

Population structure

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 [21]. 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. These initiators interact (i) in a prisoner's dilemma game with a random neighbour (i.e. individual in a neighbouring site); and (ii) in horizontal cultural transmission with a random neighbour (with replacement, i.e. possibly the same neighbour). 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 neighbour is picked for both interactions, or $\alpha = 1/M$, where M is the number of neighbours. On an infinite grid, $M = 8$ (i.e. Moore neighbourhood [29]), 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 neighbours, 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 neighbouring 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. $v = 1$).

The outcomes of stochastic simulations with such a structured population are shown in Figure 4, 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 4c. Changing the simulation so that selection is local (i.e. sites can only be settled by offspring of neighbouring parents) had only a minor

298 effect on the agreement with the derived conditions (Figure S2).

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

Discussion

304 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
306 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 2 and 3).
308 Importantly, cooperation can increase when initially rare (i.e. invade a population of defectors) if and
only if, rewriting inequality 13, $c \cdot v(1 - T_B) < b \cdot v\alpha T_A + (T_A - T_B)$, namely, the effective cost of
310 cooperation (left-hand side) is smaller than 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
312 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
314 ($T_A < T_B$) and both c and α are intermediate (yellow areas in Figures 2 and 3).

We find that stronger interaction-transmission association α leads to evolution of higher frequency
316 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
318 decreased population mean fitness (Result 2, Figure S1). With $\alpha = 0$, the benefit of cooperation cannot
facilitate its evolution; it can only succeed if horizontal transmission is biased in its favor.

320 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 , facil-
322 itates 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
324 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 cooper-
326 ation, and the rate of vertical transmission must be high enough ($v > \hat{v}$) for cooperation to fix in the
population.

328 Our deterministic model provides a good approximation to outcomes of simulations of a stochastic
model with population structure in which individuals can only interact with and transmit to their
330 neighbours. In these structured populations interaction-transmission association arises due to both
social interactions and horizontal cultural transmission being local (Figure 4).

332 Feldman et al. [16] 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
334 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
336 transmission into our model. Woodcock [19] stressed the significance of non-vertical transmission for
the evolution of cooperation and carried out simulations with prisoner's dilemma payoffs but without
338 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
340 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.

342 To understand the role of horizontal transmission, we first review the role of *assortment*. Eshel and
Cavalli-Sforza [10] showed that altruism can evolve when the tendency for *assortative meeting*, i.e.

344 for individuals to interact with others of their own phenotype, is strong enough. Fletcher and Doebeli
345 [11] further argued that a general explanation for the evolution of altruism is given by *assortment*: the
346 correlation between individuals that carry an altruistic trait and the amount of altruistic behavior in
347 their interaction group (see also Bijma and Aanen [12]). They suggested that to explain the evolution
348 of altruism, we should seek mechanisms that generate assortment, such as spatial structure, repeated
349 interactions, and individual recognition. Our results highlight another mechanism for generating
350 assortment: an association between social interactions and horizontal transmission that creates a
351 correlation between one's partner for interaction and the partner for transmission. This mechanism
352 does not require repeated interactions, spatial structure, or individual recognition. We show that
353 high levels of such interaction-transmission association greatly increase the potential for evolution of
354 cooperation. With enough interaction-transmission association, cooperation can increase in frequency
355 when initially rare even when there is horizontal transmission bias against it ($T_A < T_B$).

356 How does non-vertical transmission generate assortment? Lewin-Epstein et al. [20] and Lewin-
357 Epstein and Hadany [21] suggested that microbes that induce their hosts to act altruistically can
358 be favored by selection, which may help to explain the evolution of cooperation. From the kin
359 selection point-of-view, if microbes can be transmitted *horizontally* from one host to another during
360 host interactions, then following horizontal transmission the recipient host will carry microbes that
361 are closely related to those of the donor host, even when the two hosts are (genetically) unrelated.
362 From the assortment point-of-view, infection by behavior-determining microbes during interactions
363 effectively generates assortment because a recipient of help may be infected by a behavior-determining
364 microbe and consequently become a helper. Cultural horizontal transmission can similarly generate
365 assortment between cooperators and enhance the benefit of cooperation if cultural transmission and
366 helping interactions occur between the same individuals, i.e. when there is interaction-transmission
367 association, so that the recipient of help may also be the recipient of the cultural trait for cooperation.
368 Thus, with horizontal transmission, “assortment between focal cooperative players and cooperative
369 acts in their interaction environment” [11] is generated not because the helper is likely to be helped,
370 but rather because the helped is likely to become a helper.

Acknowledgements

372 We thank Lilach Hadany, Ayelet Shavit, and Kaleda Krebs Denton for discussions and comments. This work was
373 supported in part by the Clore Foundation Scholars Programme (OLE), the Morrison Institute for Population
374 and Resources Studies at Stanford University (MWF), Israel Science Foundation (YR 552/19), and Minerva
375 Stiftung Center for Lab Evolution (YR).

376 References

- [1] Robert Axelrod and William D Hamilton. The evolution of cooperation. *Science*, 211(4489):1390–1396,
378 1981.
- [2] Lee Alan Dugatkin. *Cooperation among Animals: An Evolutionary Perspective*. Oxford University Press
380 on Demand, 1997.
- [3] Adrian V Jaeggi and Michael Gurven. Natural cooperators: food sharing in humans and other primates.
382 *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
384 Psychology*, 55(1):123, 1962.
- [5] Peter B Stacey and Walter D Koenig, editors. *Cooperative breeding in birds: long term studies of ecology
386 and behaviour*. Cambridge University Press, 1990.
- [6] Indrikis Krams, Tatjana Krama, Kristine Igaune, and Raivo Mänd. Experimental evidence of reciprocal
388 altruism in the pied flycatcher. *Behavioral Ecology and Sociobiology*, 62(4):599–605, 2008.

- 390 [7] Barry Sinervo, Alexis Chaine, Jean Clobert, Ryan Calsbeek, Lisa Hazard, Lesley Lancaster, Andrew G
McAdam, Suzanne Alonso, 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.
- 394 [8] J. B. S. Haldane. *The Causes of Evolution*. Longmans, London, 1932.
- 396 [9] William D Hamilton. The genetical evolution of social behaviour. ii. *Journal of Theoretical Biology*, 7
(1):17–52, 1964.
- 398 [10] 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.
- 400 [11] 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.
- 402 [12] Piter Bijma and Duur K. Aanen. Assortment, Hamilton’s rule and multilevel selection. *Proc. R. Soc. B
Biol. Sci.*, 277(1682):673–675, 2010.
- 404 [13] 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.
- 406 [14] Luigi Luca Cavalli-Sforza and Marcus W Feldman. *Cultural transmission and evolution: A quantitative
approach*. Number 16. Princeton University Press, 1981.
- 408 [15] Peter J Richerson and Robert Boyd. *Not by Genes Alone: How Culture Transformed Human Evolution*.
University of Chicago Press, 2008.
- 410 [16] 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.
- 412 [17] Stephen J Lycett and John AJ Gowlett. On questions surrounding the acheulean ‘tradition’. *World
Archaeology*, 40(3):295–315, 2008.
- 414 [18] 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.
- 416 [19] Scott Woodcock. The significance of non-vertical transmission of phenotype for the evolution of altruism.
Biology and Philosophy, 21(2):213–234, 2006.
- 418 [20] Ohad Lewin-Epstein, Ranit Aharonov, and Lilach Hadany. Microbes can help explain the evolution of
host altruism. *Nature Communications*, 8:14040, 2017.
- 420 [21] 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.
- 422 [22] 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.
- 424 [23] Matthew R. Zefferman. Mothers teach daughters because daughters teach granddaughters: the evolution
of sex-biased transmission. *Behav. Ecol.*, 27(4):1172–1181, 2016.
- 426 [24] Marcus W. Feldman. Selection for linkage modification: I. Random mating populations. *Theor. Popul.
Biol.*, 3:324–346, 1972.
- 428 [25] Uri Liberman and Marcus W. Feldman. A general reduction principle for genetic modifiers of recombina-
tion. *Theor. Popul. Biol.*, 30(3):341–71, dec 1986.
- 430 [26] Uri Liberman and Marcus W. Feldman. Modifiers of mutation rate: A general reduction principle. *Theor.
Popul. Biol.*, 30:125–142, 1986.

- 432 [27] Uri Liberman. External stability and ESS: criteria for initial increase of new mutant allele. *J. Math. Biol.*,
 26(4):477–485, 1988.
- 434 [28] 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.
- 436 [29] 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.
- 438 [30] 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.
- 440 [31] 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.

442 **Tables**

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	$[0, 1]$
\hat{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	$(0, 1)$

Figures

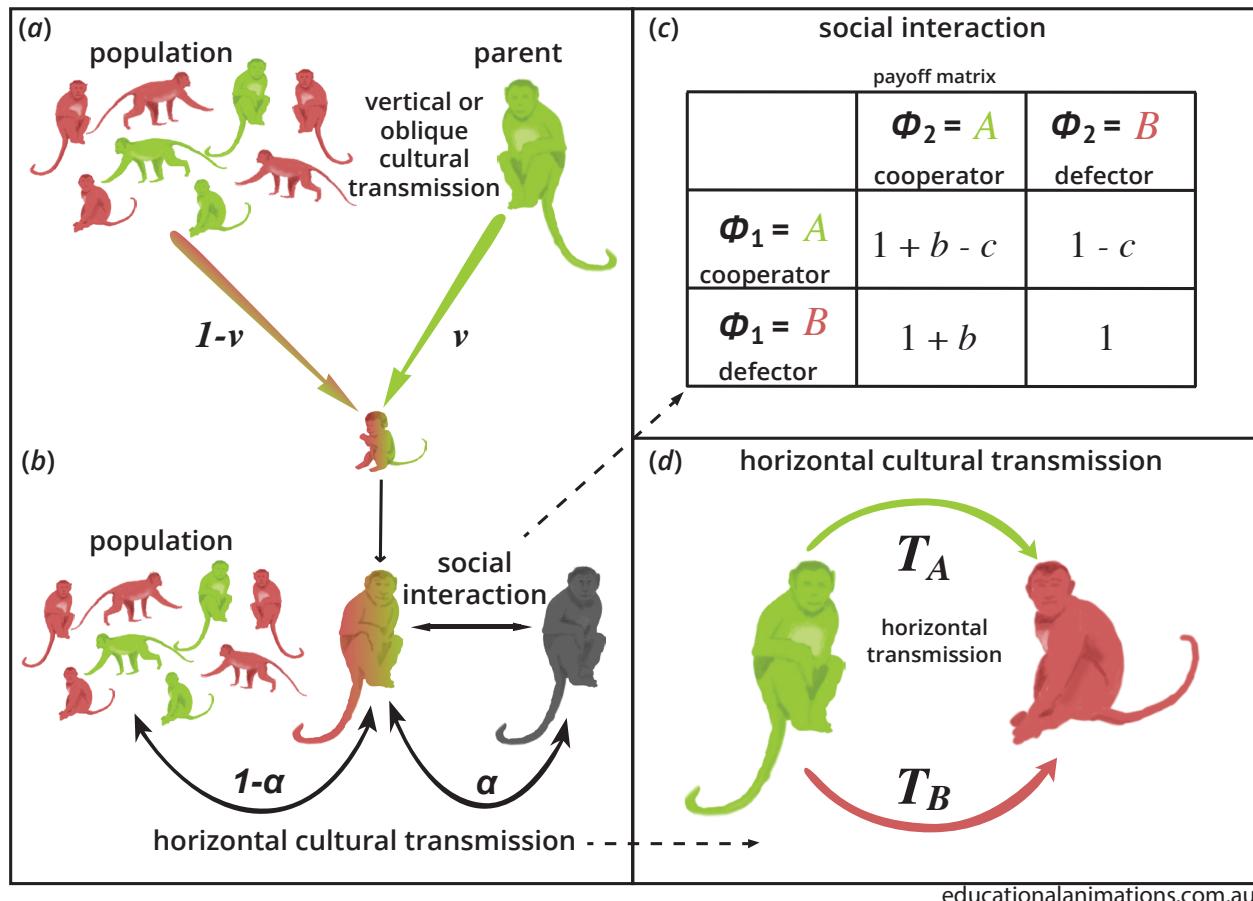


Figure 1: Model illustration. (a) First, offspring inherit their parent's phenotype via vertical cultural transmission with probability v , or the phenotype of a random non-parental adult via oblique cultural transmission with probability $1 - v$. (b) Second, adults socially interact in pairs in a prisoner's 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.

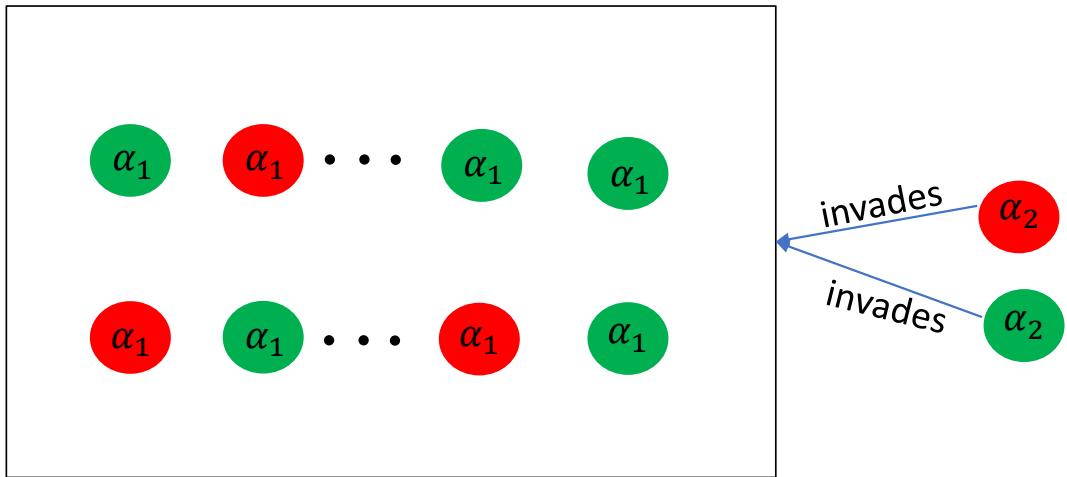


Figure 2: Evolution of cooperation under vertical, oblique, and horizontal cultural transmission. The figure shows parameter ranges for global fixation of cooperation (green), global fixation of defection (red), fixation of either cooperation or defection depending on the initial conditions, i.e. unstable 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 γ_1 and γ_2 (Eqs. 12) 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. 17) 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)** $b = 2$, $T_B = 0.7$, $c = 0.5$.

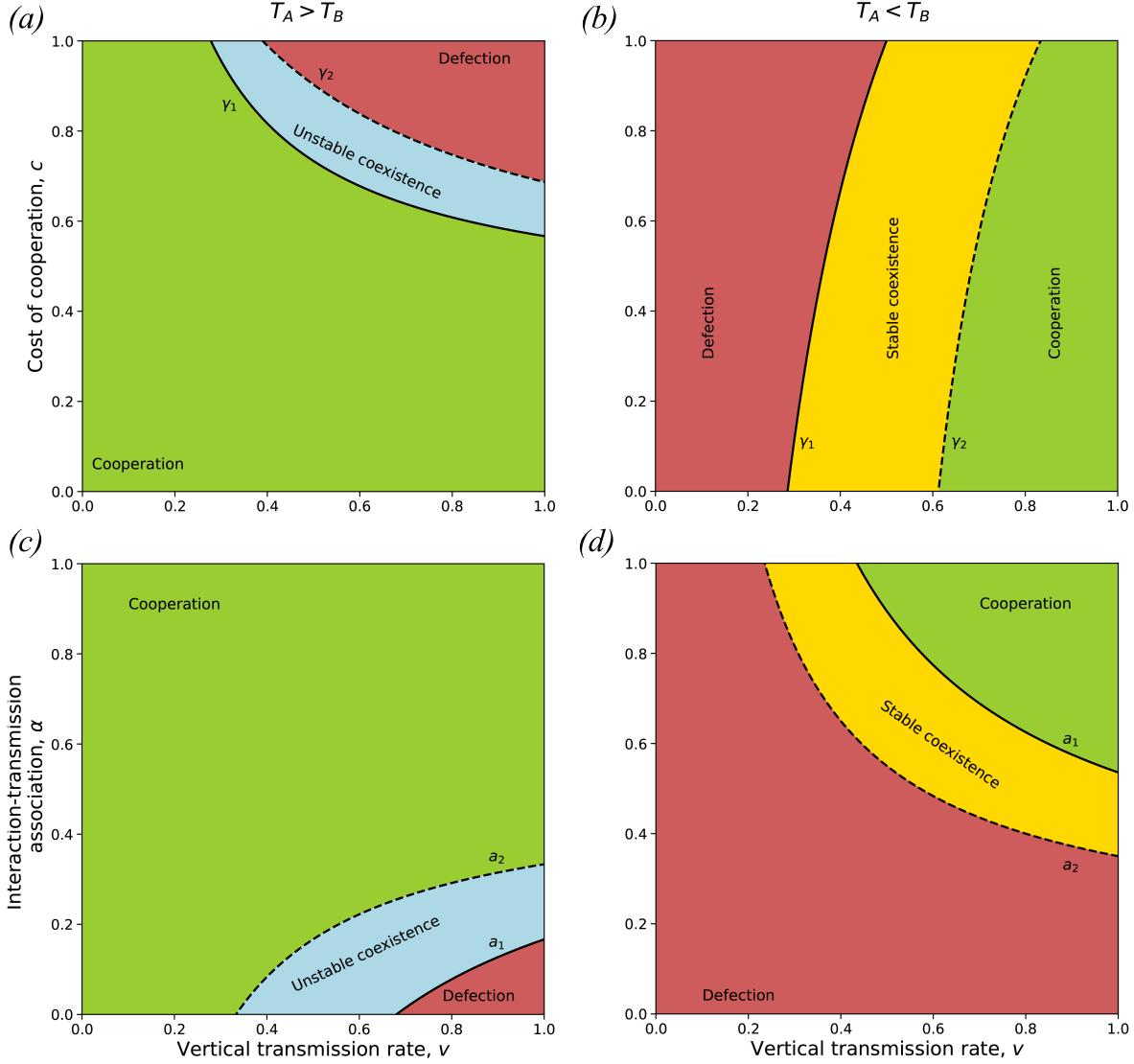


Figure 3: Evolution of cooperation under vertical and horizontal cultural transmission ($v=1$).

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 γ_1 and γ_2 (Eq. 12) 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. 17) 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. 10. 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.

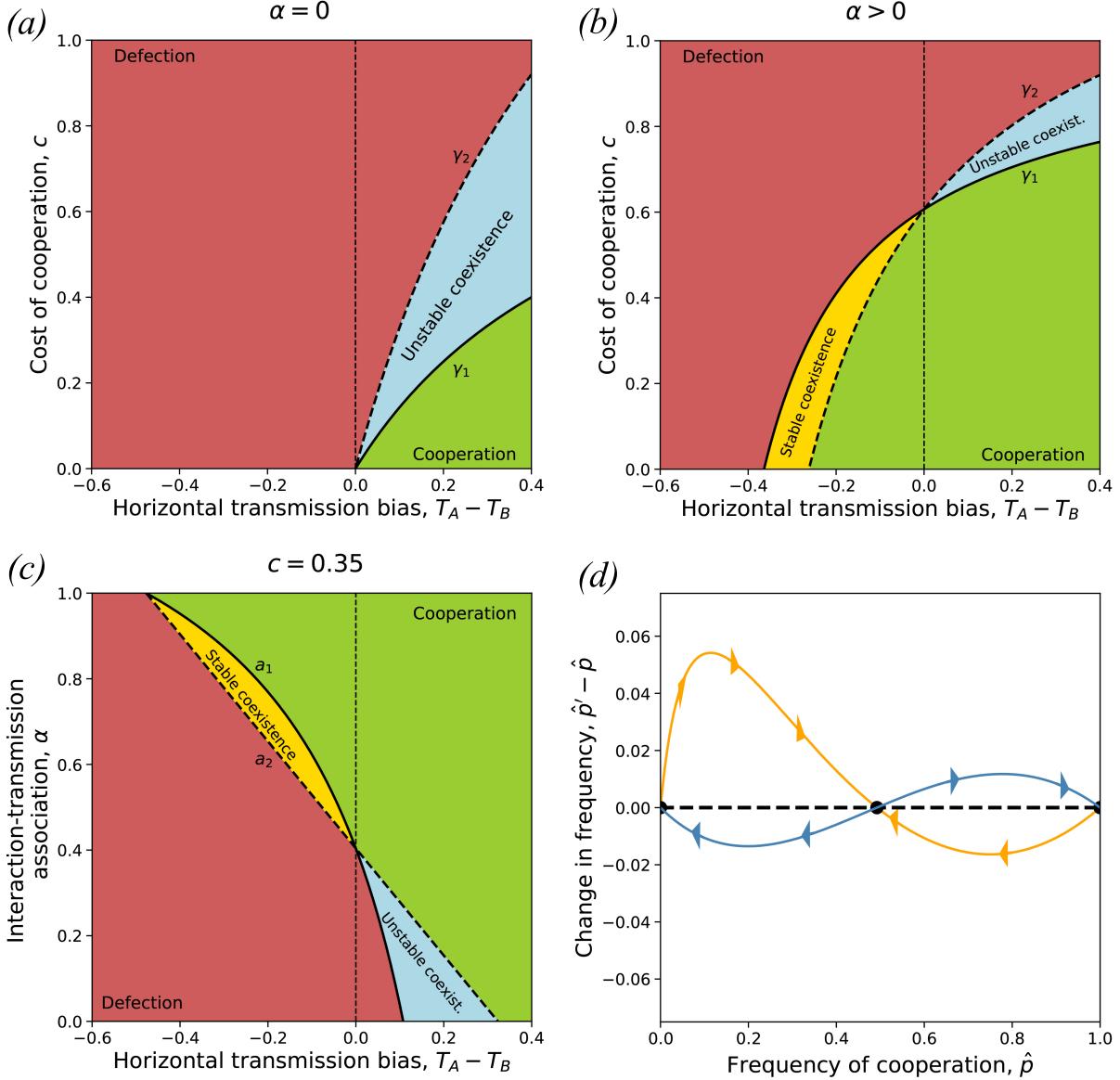


Figure 4: Evolution of cooperation in a structured population. (a-b) The expected frequency of cooperators in a structured population after 10,000 generations is shown (red for 0%, green for 100%) as a function of both the cost of cooperation, c , on the y-axis, and either the symmetric horizontal transmission 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 14 for panel (a) and in Eqs. 12 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 [30]. 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 $c = 0.02$.

444 Appendices

Appendix A Local stability criterion

446 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

$$448 \quad p' < p \Leftrightarrow f(p)/p < 0 \Leftrightarrow \frac{f'(0) \cdot p + O(p^2)}{p} < 0 \Leftrightarrow f'(0) + O(p) < 0. \quad (\text{A1})$$

Therefore, by definition of big-O notation, if $f'(0) < 0$ then there exists $\epsilon > 0$ such that for any local perturbation $450 \quad 0 < p < \epsilon$, 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

$$452 \quad 1 - p' < 1 - p \Leftrightarrow -\frac{f(p)}{1 - p} < 0 \Leftrightarrow \frac{f'(1)(p - 1) + O((p - 1)^2)}{p - 1} < 0 \Leftrightarrow f'(1) - O(1 - p) < 0. \quad (\text{A2})$$

Therefore, if $f'(1) < 0$ then there exists $\epsilon > 0$ such that for any $1 - \epsilon < 1 - p < 1$ we have $1 - p' < 1 - p$; that $454 \quad$ is, p' is closer to one than p .

Appendix B Equilibria and stability

456 Let $f(\hat{p}) = \bar{w}(\hat{p}' - \hat{p})$. Then, using SymPy [31], a Python library for symbolic mathematics, this simplifies to

$$458 \quad f(\hat{p}) = \bar{w}(\hat{p}' - \hat{p}) = \beta_1 \hat{p}^3 + \beta_2 \hat{p}^2 + \beta_3 \hat{p}, \quad (\text{B1})$$

where

$$460 \quad \begin{aligned} \beta_1 &= [c(1 - v) - b(1 - \alpha v)](T_A - T_B), \\ \beta_2 &= -\beta_1 - \beta_3, \\ \beta_3 &= \alpha b v T_A - c v (1 - T_B) + (T_A - T_B). \end{aligned} \quad (\text{B2})$$

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,

$$462 \quad f(\hat{p}) = \hat{p}(1 - \hat{p})[\alpha b v T - c v (1 - T)]. \quad (\text{B3})$$

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 that

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

466 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

$$468 \quad \hat{p}^* = \frac{\beta_3}{\beta_1} = \frac{\alpha b v T_A - c v (1 - T_B) + (T_A - T_B)}{[c(1 - v) - b(1 - \alpha v)](T_A - T_B)}. \quad (\text{B5})$$

Note that the sign of the cubic (Eq. B1) at positive (negative) infinity is equal (opposite) to the sign of β_1 . If $470 \quad T_A > T_B$, then

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

472 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.

478 Similarly, if $T_A < T_B$, then

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

480 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.

486 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. *Stable polymorphism with 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 $\beta_1 < \beta_3 < 0$.

We now proceed to use the cost thresholds, γ_1 and γ_2 , and the vertical transmission threshold, \hat{v} (Eq. 12).

494 First, assume $T_A < T_B$. $\beta_3 < 0$ requires $\gamma_1 < c$. For $\beta_3 < \beta_1$ we need $c[v(1 - T_B) + (1 - v)(T_A - T_B)] > b\alpha v T_B + (1 + b)(T_A - T_B)$. Note that the expression in the square brackets is positive if and only if $v > \hat{v}$. Thus, 496 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 498 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 $c[v(1 - T_B) + (1 - v)(T_A - T_B)] < b\alpha v T_B + (1 + b)(T_A - T_B)$. 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 $c > \gamma_2$.

502 This analysis is summarized in Result 1.

Appendix C Effect of interaction-transmission association on mean fitness

504 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}^* ,

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

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

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

510 Therefore, the numerator in Eq. C1 is positive. Since $T_A < T_B$, the denominator in Eq. C1 is also positive, and hence the derivative $\partial \hat{p}^* / \partial \alpha$ is positive. Thus, the population mean fitness increases as interaction-transmission 512 association α increases.

Appendix D Reduction principle

514 We assume here that $v = 1$, i.e. no oblique transmission, and therefore $\hat{p} = \dot{p}$. Denote the frequencies of the
 516 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,

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

518 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 population mean fitness.

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

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

522 setting $\alpha = \alpha_1$ and $v = 1$ in Eq. 11. When $v = 1$, \dot{p}^* is a feasible polymorphism ($0 < \dot{p}^* < 1$) if $T_A < T_B$ and $\gamma_2 < c < \gamma_1$ (Result 1).

524 The local stability of \mathbf{p}^* to the introduction of allele m is determined by the linear approximation \mathbf{L}^* of the transformation in Eq. D1 near \mathbf{p}^* (i.e. the Jacobian of the transformation at the equilibrium). \mathbf{L}^* is known
 526 to have a block structure, with the diagonal blocks occupied by the matrices \mathbf{L}_{in}^* and \mathbf{L}_{ex}^* [26, 28]. The latter is the external stability matrix: the linear approximation to the transformation near \mathbf{p}^* involving only the
 528 pheno-genotypes Am and Bm , derived from Eq. D1, with $\bar{w}^* = 1 + (b - c)\dot{p}^*$ as the stable population mean fitness,

$$\begin{aligned}\mathbf{L}_{ex}^* &= \frac{1}{\bar{w}^*} \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix} = \frac{1}{\bar{w}^*} \begin{bmatrix} \frac{\partial \bar{w} \dot{p}'_3}{\partial \dot{p}_3}(\mathbf{p}^*) & \frac{\partial \bar{w} \dot{p}'_3}{\partial \dot{p}_4}(\mathbf{p}^*) \\ \frac{\partial \bar{w} \dot{p}'_4}{\partial \dot{p}_3}(\mathbf{p}^*) & \frac{\partial \bar{w} \dot{p}'_4}{\partial \dot{p}_4}(\mathbf{p}^*) \end{bmatrix} = \\ 530 & \frac{1}{\bar{w}^*} \begin{bmatrix} (1+b\dot{p}^*-c)(1-T_B(1-\dot{p}^*))+b\dot{p}^*\alpha_2T_B(1-\dot{p}^*) & (1+b\dot{p}^*)T_A\dot{p}^*+b\dot{p}^*\alpha_2T_A(1-\dot{p}^*) \\ (1+b\dot{p}^*-c)T_B(1-\dot{p}^*)-b\dot{p}^*\alpha_2T_B(1-\dot{p}^*) & (1+b\dot{p}^*)(1-T_AP^*)-b\dot{p}^*\alpha_2T_A(1-\dot{p}^*) \end{bmatrix}.\end{aligned}\tag{D3}$$

Because we assume that \mathbf{p}^* is internally stable (i.e. locally stable to small perturbations in the frequencies
 532 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
 534 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
 536 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 multiplication by a positive factor doesn't change the sign, and using the properties of the determinant, we have

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

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

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

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

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

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

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

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

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

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 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 $R'(1)$. However, we have numerically validated that the leading eigenvalue is greater than one if and only if $\alpha_1 > \alpha_2$.