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

Dor Cohen[1], Ohad Lewin-Epstein[2], Marcus W. Feldman[3], and Yoav Ram[1,4,5,*]

[1]School of Computer Science, Interdisciplinary Center Herzliya, Herzliya, Israel
[2]School of Plant Sciences and Food Security, Tel Aviv University, Tel Aviv, Israel
[3]Department of Biology, Stanford University, Stanford, CA
[4]School of Zoology, Tel Aviv University, Tel Aviv, Israel
[5]Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel
[*]Corresponding author: yoav@yoavram.com

November 14, 2020

Abstract

We study the cultural evolution of cooperation under vertical and non-vertical cultural transmission. Conditions are found for fixation and coexistence of cooperation and defection. The evolution of cooperation is facilitated by 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. When it does, selection will reduce the association between social interactions and horizontal transmission, which leads to a decreased frequency of cooperation and lower population mean fitness. The derived conditions are compared to outcomes of stochastic simulations of structured populations. Comparisons are drawn with Hamilton's rule and the concepts of relatedness and assortment.

Introduction

Cooperative behavior can reduce an individual's fitness and increase the fitness of its conspecifics or competitors (**?**). Nevertheless, cooperative behavior appears to occur in many non-human animals (**?**), including primates (**?**), rats (**?**), birds (**??**), and lizards (**?**). Evolution of cooperative behavior remains an important conundrum in evolutionary biology (**?**, Appendix).

Since the work of **?** and **?**, theories for the evolution of cooperative and altruistic behaviors have been intertwined often under the rubric of *kin selection*. Kin selection theory posits that natural selection is more likely to favor cooperation between more closely related individuals. The importance of *relatedness* to the evolution of cooperation and altruism was demonstrated by **?**, 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*:

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

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

? studied a related model for the evolution of cooperative behavior. Their model included *assortative meeting*, or non-random encounters, where a fraction m of individuals in the population each interact 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[1] (**?**, eq. 3.2)

$$c < b \cdot m, \tag{2}$$

where b and c are the benefit and cost of cooperation.

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

$$c < b \cdot (p_C - p_D), \tag{3}$$

where p_C is the probability that a cooperator receives help, and p_D is the probability that a defector receives help.[2] **?** gave a result related to inequality **??** for other types of games.

In this paper we study the evolution of a cooperative behavior that is subject to *cultural transmission*, which allows an individual to acquire attitudes or behavioral traits from other individuals in its social group through imitation, learning, or other modes of communication (**??**). **?** introduced the first model for the evolution of altruism by cultural transmission and demonstrated that if the fidelity of cultural transmission of altruism is φ, then the condition for evolution of altruism in the case of sib-to-sib altruism is (**?**, Eq. 16)

$$c < b \cdot \varphi - \frac{1 - \varphi}{\varphi}. \tag{4}$$

In inequality **??**, φ takes the role of relatedness (r in inequality **??**) or assortment (m in inequality **??**), but the effective benefit $b \cdot \varphi$ is reduced by $(1 - \varphi)/\varphi$. This shows that under a combination of genetic

---

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

[2]Inequality **??** generalizes inequalities **??** and **??** 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.

and cultural transmission, the condition for the evolutionary success of altruism entails a modification
of Hamilton's rule (**??**).

Cultural transmission may be viewed as vertical, horizontal or oblique: vertical transmission occurs
between parents and offspring, horizontal transmission occurs between individuals from the same
generation, and oblique transmission occurs to offspring from the generation to which their parents
belong (i.e. from non-parental adults). Evolution under either of these transmission models can be be
more rapid than under pure vertical transmission (**???**). Both **?** and **?** 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 microbes that
manipulate their host's behavior.) Some of the analyses by **?** can be applied to cultural transmission,
because models of cultural transmission are mathematically similar to those for transmission of
infectious diseases (**?**).

Here, we study cultural-evolution models of cooperation that include both vertical and non-vertical
transmission. In our models behavioral changes are mediated by cultural transmission that can occur
specifically during social interactions. For instance, there may be an association between the choice of
partner for social interaction and the choice of partner for cultural transmission. As another example,
when an individual interacts with an individual of a different phenotype, exposure to the latter may lead
the former to convert its phenotype. Our results demonstrate that cultural transmission can enhance
the evolution of cooperation even when genetic transmission cannot, partly because it facilitates the
generation of assortment (**?**), and partly because non-vertical transmission can diminish the effect of
natural selection (**?**). This further emphasizes that treatment of cooperation as a cultural trait, rather
than a genetic one, can lead to a broader understanding of its evolutionary dynamics.

Models

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 with
probability v or from a random individual in the parental population via oblique transmission with
probability $(1 - v)$. Following **?**, given that the parent's phenotype is ϕ and assuming uni-parental
inheritance (**?**), the conditional probability that the phenotype ϕ' of the offspring is A is

$$P(\phi' = A \mid \phi) = \begin{cases} v + (1 - v)p, & \text{if } \phi = A \\ (1 - v)p, & \text{if } \phi = B \end{cases}, \tag{5}$$

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

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

$$\begin{aligned} \hat{p} &= \tilde{p}[v + (1 - v)p] + (1 - \tilde{p})[(1 - v)p] = \\ & v\tilde{p} + (1 - v)p . \end{aligned} \tag{6}$$

Individuals are assumed to interact according to a *prisoner's dilemma*. Specifically, individuals
interact in pairs; a cooperator suffers a fitness cost $0 < c < 1$, and its partner gains a fitness benefit b,
where we assume $c < b$. **??** shows the payoff matrix, i.e. the fitness of an individual with phenotype
ϕ_1 when interacting with a partner of phenotype ϕ_2.

Social interactions occur randomly: two juvenile individuals with phenotype A interact with proba-
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 $2\hat{p}(1 - \hat{p})$.

102 Horizontal cultural transmission occurs between pairs of individuals from the same generation. It occurs between socially interacting partners with probability α, or between a random pair with

104 probability $1 - \alpha$ (see **??**). However, horizontal transmission is not always successful, as one partner may reject the other's phenotype. The probability of successful horizontal transmission of phenotypes

106 A and B are T_A and T_B, respectively (**??**).

Therefore, the frequency p' of phenotype A among adults in the next generation, after horizontal

108 transmission, is

$$
\begin{aligned}
p' =& \hat{p}^2\big[\alpha + (1 - \alpha)\big(\hat{p} + (1 - \hat{p})(1 - T_B)\big)\big] + \\
& \hat{p}(1 - \hat{p})\big[\alpha(1 - T_B) + (1 - \alpha)\big(\hat{p} + (1 - \hat{p})(1 - T_B)\big)\big] + \\
& (1 - \hat{p})\hat{p}\big[\alpha T_A + (1 - \alpha)\hat{p}T_A\big] + \\
& (1 - \hat{p})^2\big[(1 - \alpha)\hat{p}T_A\big] \, ,
\end{aligned}
\tag{7}
$$

110 which simplifies to

$$
p' = \hat{p}^2(T_B - T_A) + \hat{p}(1 + T_A - T_B) \, .
\tag{8}
$$

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

114

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

where fitness values are taken from **??** and **??**, and the population mean fitness is $\bar{w} = 1 + \hat{p}(b - c)$.

116 **??** can be simplified to

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

118 Starting from **??** with $\hat{p}' = v\tilde{p}' + (1 - v)p'$, we substitute **??** for p' and **??** for $\tilde{p}'$ and obtain

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

120 **??** lists the model variables and parameters.

Results

In the following sections, we determine the equilibria of the model in **??**, namely, solutions of $\hat{p}' = \hat{p}$, and analyze their local stability. We then analyze the evolution of a modifier of social association. Finally, we compare derived conditions to outcomes of stochastic simulations with a structured population.

Evolution of cooperation

The equilibria of our model, i.e. the fixed points of **??**, are $\hat{p} = 0$, $\hat{p} = 1$, and (see **??**)

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

(12)

To find conditions on the cost of cooperation, c, define the following cost thresholds, γ_1 and γ_2, and the vertical transmission threshold, $\hat{v}$,

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

(13)

The full analysis is given in Appendix **??**. The following result summarizes the possible outcomes.

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

1. *Fixation of cooperation: if* (i) $T_A \geq T_B$ *and* $c < \gamma_1$; *or if* (ii) $T_A < T_B$ *and* $v > \hat{v}$ *and* $c < \gamma_2$.

2. *Fixation of defection: if* (iii) $T_A \geq T_B$ *and* $\gamma_2 < c$; *or if* (iv) $T_A < T_B$ *and* $\gamma_1 < c$.

3. *Stable polymorphism: if* (v) $T_A < T_B$ *and* $v < \hat{v}$ *and* $c < \gamma_1$; *or if* (vi) $T_A < T_B$ *and* $v > \hat{v}$ *and* $\gamma_2 < c < \gamma_1$.

4. *Unstable polymorphism: if* (vii) $T_A > T_B$ *and* $\gamma_1 < c < \gamma_2$.

These conditions are illustrated in **??**a and **??**b.

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

Remark 1 (Condition for cooperation to increase from rarity: cost of cooperation). *If the initial frequency of cooperation is very close to zero, then its frequency will increase if the cost of cooperation is low enough,*

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

(14)

This unites the conditions for fixation of cooperation and for stable polymorphism, both of which lead to instability of the fixation state associated with defection, $\hat{p} = 0$.

Importantly, increasing social association α increases the cost threshold ($\partial\gamma_1/\partial\alpha > 0$), making it easier for cooperation to increase from rarity. Similarly, increasing the horizontal transmission of cooperation, T_A, increases the threshold ($\partial\gamma_1/\partial T_A > 0$), facilitating the evolution of cooperation. However, increasing the horizontal transmission of defection, T_B, can increase or decrease the cost threshold, but it increases the cost threshold when it is already above one ($c < 1 < \gamma_1$): $\partial\gamma_1/\partial T_B$

is positive when $T_A > \frac{1}{1+\alpha b v}$, which gives $\gamma_1 > 1/v$. Therefore, increasing T_B decreases the cost threshold and limits the evolution of cooperation, but only if $T_A < \frac{1}{1+\alpha b v}$.

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 get $\partial\gamma_1/\partial v > 0$, and therefore as the vertical transmission rate increases, the cost threshold increases, making it easier for cooperation to evolve from rarity. 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 evolve from rarity.

In general, this condition cannot be formulated in the form of Hamilton's rule due to the bias in horizontal transmission, represented by $T_A - T_B$. When there is no horizontal transmission bias, $T_A = T_B$, the next remark applies.

Remark 2 (Unbiased horizontal transmission). *If horizontal transmission is unbiased, $T = T_A = T_B$, then cooperation will take over the population from any initial frequency if the cost is low enough,*

$$c < b \cdot \frac{v\alpha T}{1 - T} \ . \tag{15}$$

This condition is obtained by setting $T = T_A = T_B$ in inequality (**??**) and can be interpreted as a version of Hamilton's rule ($c < b \cdot r$, inequality **??**) or as a version of inequality (**??**), where $v\alpha T/(1 - T)$ can be regarded as the *effective relatedness* or *effective assortment*, respectively. **??**a illustrates this condition for $v = 1$. Note that the right-hand side equals γ_1 when $T = T_A = T_B$.

In the next remarks we examine the effect of social association on the evolution of cooperation.

Remark 3 (No social association). *Without social association ($\alpha = 0$), cooperation will increase when it is rare if there is horizontal transmission bias for cooperation, $T_A > T_B$, and*

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

??a illustrates this condition, which is obtained by setting $\alpha = 0$ in inequality (**??**). Importantly, the benefit of cooperation, b, does not affect the evolution of cooperation in the absence of social association, and the outcome is determined only by cultural transmission.

Remark 4 (Perfect social association). *With perfect social association ($\alpha = 1$), cooperation will increase when it is rare if*

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

This condition is obtained by setting $\alpha = 1$ in inequality (**??**).

In the absence of oblique transmission, $v = 1$, the only equilibria are the fixation states, $\tilde{p} = 0$ and $\tilde{p} = 1$, and cooperation will evolve from any initial frequency (i.e., $\tilde{p}' > \tilde{p}$) if inequality **??** applies. This is similar to case of microbe-associated cooperation studied by **?**, and therefore when $v = 1$, this remark is equivalent to their eq. (1).

In light of Remarks **??** and **??**, it is interesting to examine the general effect of social association on the evolution of cooperation. Define the social association thresholds, a_1 and a_2, as

$$a_1 = \frac{c \cdot v(1 - T_A) - (T_A - T_B)(1 + b - c)}{b \cdot v \cdot T_B}, \quad a_2 = \frac{c \cdot v(1 - T_B) - (T_A - T_B)}{b \cdot v \cdot T_A} \ . \tag{18}$$

196 Then the following applies.

Remark 5 (Condition for cooperation to increase from rarity: social association). *Cooperation will increase when rare if social association is high enough, specifically if $a_2 < \alpha$.*

200 Figures **??**c, **??**c, and **??**d 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 positive x-axis).
202 With horizontal bias favoring defection, $T_A < T_B$, cooperation can fix from any frequency if social association is high, $a_1 < \alpha$ (green area with $T_A < T_B$), and can also increase when rare and reach
204 stable polymorphism with defection if social association is intermediate, $a_2 < \alpha$ (yellow area). Without horizontal bias, $T_A = T_B$, fixation of cooperation occurs if social association is high enough,
206 $\frac{c}{b \cdot v} \cdot \frac{1-T}{T} < \alpha$ (inequality **??**; in this case $a_1 = a_2$).

Interestingly, because $\text{sign}(\partial a_2 / \partial v) = \text{sign}(T_A - T_B)$, the effect of the vertical transmission rate v on
208 a_1 and a_2 depends on the horizontal transmission bias. That is, with horizontal bias for cooperation, $T_A > T_B$, evolution of cooperation is facilitated by oblique transmission, whereas with horizontal
210 bias for defection, $T_A < T_B$, evolution of cooperation is facilitated by vertical transmission. This is demonstrated in Figures **??**c and **??**d.

212 The next remarks explain the roles of vertical and oblique transmission in the evolution of cooperation.

214 **Remark 6** (Necessary condition for fixation of cooperation: vertical transmission rate). *Fixation of cooperation is possible only if the vertical transmission rate is high enough,*

$$v > \hat{v} = \frac{T_B - T_A}{1 - T_A} \ . \tag{19}$$

This condition does not guarantee fixation of cooperation; rather, if this condition does not apply then
218 cooperation cannot fix. If horizontal transmission is biased for cooperation, $T_A > T_B$, cooperation can fix with any vertical transmission rate (because $\hat{v} < 0$). In contrast, if horizontal transmission is
220 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 **??**b and **??**d).

222 With only vertical transmission ($v = 1$), the following applies.

224 **Remark 7** (Condition for cooperation to increase from rarity: vertical and horizontal transmission). *In the absence of oblique transmission ($v = 1$) cooperation will increase when rare if*

$$c < \frac{b\alpha T_A + (T_A - T_B)}{1 - T_B} \ , \tag{20}$$

which can also be written as

$$\frac{c(1 - T_B) - (T_A - T_B)}{bT_A} < \alpha \ . \tag{21}$$

These conditions are obtained by setting $v = 1$ in inequalities (**??**) and (**??**).

230 In the absence of vertical transmission ($v = 0$) we have the following.

232 **Remark 8** (Condition for global fixation of cooperation: oblique and horizontal transmission). *In the absence of vertical transmission ($v = 0$), the frequency of the cooperator phenotype among*
234 *adults increases every generation, i.e. $p' > p$, if there is a horizontal transmission bias in favor of cooperation, namely*

$$T_A > T_B \ . \tag{22}$$

7

This condition is found by setting $v = 0$ in **??**, which becomes $\hat{p} = p$, and then **??** becomes $p' = p^2(T_B - T_A) + p(1 + T_A - T_B)$. That is, in the absence of vertical transmission, selection plays no role in the evolution of cooperation (i.e. b and c do not affect p'). The dynamics are determined solely by differential horizontal transmission of the two phenotypes, namely, the relative tendency of each phenotype to be horizontally transmitted to peers. Note that with no bias in horizontal transmission, $T_A = T_B$, phenotype frequencies are static, $p' = p$.

Finally, the last remark focuses on the polymorphism of cooperation and defection.

Remark 9 (Polymorphism of cooperation and defection). *Cooperation and defection can coexist at frequencies $\tilde{p}^*$ and $1 - \tilde{p}^*$ (**??**). When it is feasible, this equilibrium is stable if there is horizontal transmission bias for defection, $T_A < T_B$, and unstable if there is horizontal bias for cooperation, $T_A > T_B$.*

The yellow and blue areas in Figures **??** and **??** show cases of stable and unstable polymorphism, respectively. When $\tilde{p}^*$ is unstable, cooperation will fix if its initial frequency is $p > \tilde{p}^*$, and defection will fix $p < \tilde{p}^*$; this occurs when there is horizontal transmission bias for cooperation, $T_A > T_B$, and the cost is intermediate, $\gamma_1 < c < \gamma_2$. **??**a shows the mapping $\tilde{p} \to \tilde{p}'$.

Evolution of social association

We now focus on the evolution of social 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 $\hat{p}^*$ (**??**). Note that for stable polymorphism to occur, there must be horizontal bias for defection, $T_A < T_B$, and an intermediate cost of cooperation, $\gamma_2 < c < \gamma_1$ (**??**)

The equilibrium population mean fitness is $\bar{w}^* = 1 + \hat{p}^*(b - c)$, which is increasing in $\hat{p}^*$, and $\hat{p}^*$ is increasing in α (Appendix **??**). Therefore, if social association increases, the population mean fitness also increases. But can an allele that increases social association evolve?

To answer this question, we extend our model to include a "modifier locus" (**??**) that determines social association, but has no direct effect on fitness. The modifier locus has two alleles, M and m, which induce social associations α_1 and α_2, respectively. Suppose that the population has evolved to a stable equilibrium $\hat{p}^*$ where 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", see **?**. The full analysis is given in Appendix **??**, and the following result summarizes the condition for this.

Result 2 (Reduction principle for social association). *From a stable polymorphism between cooperation and defection, a modifier allele can successfully invade the population if it decreases the social association α.*

Note that this reduction principle entails that successful invasions will reduce the frequency of cooperation, as well as the population mean fitness (**??**). Furthermore, if we assume that modifier alleles with decreased social association appear and invade the population from time to time, then, the social association will continue to decrease, further reducing the frequency of cooperation and the population mean fitness. This evolution will proceed as long as stable polymorphism occurs, that is, as long as $a_2 < \alpha < a_1$ (Remark **??**, **??**c). Thus, we can expect social association to eventually drop to a_2, the frequency of cooperation to drop to zero, and the population mean fitness to drop to one.

Population structure

280 Social 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

282 the model of **?**. 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

284 probability. In each generation, half of the individuals are randomly chosen to "initiate" interactions. Initiators then interact with a random neighbor (i.e. individual in a neighboring site) in a prisoners'

286 dilemma game (**??**) and a random neighbor (with replacement) for horizontal cultural transmission (**??**). The expected number of each of these interactions per individual per generation is one. The

288 effective social association α in this model is the probability that the same neighbor is picked for both interactions, or $\alpha = 1/m$, where m is the number of neighbors. On an infinite grid, $m = 8$, but on a

290 finite grid m can be lower in edge neighborhoods close to the grid border. As before, T_A and T_B are the probability for successful horizontal transmission of phenotypes A and B, respectively.

292 The order 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 interactions

294 of that individual with cooperative neighbors, and n_c is the total number of interactions that that individual had ($n_b \leq n_c$). Then, a new generation is generated, and sites can be settled by offspring of

296 any parent, not just neighbor parents. Thus, selection is global, rather than local, in accordance with our deterministic model. The parent is randomly drawn with probability proportional to its fitness,

298 divided by the average fitness of all potential parents. Offspring then have the same phenotype as their parents (i.e. $v = 1$).

300 The outcomes of stochastic simulations with a structured population are shown in **??**, which demonstrates that the highest cost of cooperation c that permits the evolution of cooperation agrees with the

302 conditions derived above for a model without population structure or stochasticity. An example of a case with stable polymorphism is shown in **??**. Changing the simulation so that the population is

304 structured to allow local selection (i.e. sites could only be settled by offspring of neighboring parents) had only a minor effect on the agreement with the derived conditions (**??**).

306 These comparisons between the deterministic unstructured model and the stochastic structured model show that the derived conditions can be useful for predicting the dynamics under complex scenarios.

308 Moreover, our structured populations demonstrate how our parameter for social association (α) can represent local interactions between individuals.

Discussion

310

Under a combination of vertical, oblique, and horizontal transmission with payoffs in the form of

312 payoffs from a prisoner's dilemma game, cooperation or defection can either fix or coexist, depending on the relationship between the cost and benefit of cooperation, the horizontal transmission bias,

314 and the association between social interaction and horizontal transmission (Result **??**). Importantly, cooperation can increase when initially rare (i.e. invade a population of defectors) if and only if

316 (inequality **??**),

$$c \cdot v(1 - T_B) < b \cdot v\alpha T_A + (T_A - T_B) \, , \tag{23}$$

318 namely, if the effective cost of cooperation (left-hand side) is smaller then the effective benefit plus the horizontal transmission bias (right-hand side). This condition cannot be formulated in the form of

320 Hamilton's rule, $c < b \cdot r$, due to the effect of biased horizontal transmission, represented by $(T_A - T_B)$. Remarkably, a polymorphism of cooperation and defection can be stable if horizontal transmission is

322 biased for defection ($T_A < T_B$) and both the cost of cooperation and social association are intermediate (yellow areas in Figures **??** and **??**).

9

We find that stronger social association α leads to evolution of higher frequency of cooperation and increased population mean fitness. Nevertheless, when cooperation and defection coexist, social association is expected to be reduced by natural selection, leading to extinction of cooperation and decreased population mean fitness (Result **??**). Without social association, the benefit of cooperation cannot facilitate its evolution; cooperation can only succeed if horizontal transmission is biased in its favor.

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

The conditions derived from our deterministic model provide a good approximation to outcomes of simulations of a complex stochastic model with population structure in which individuals can only interact with and transmit to their neighbors. In these structured populations social association arises due to both social interactions and horizontal cultural transmission being local.

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

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

How does non-vertical transmission generate assortment? **?** and **?** suggested that microbes that manipulate their hosts to act altruistically can 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 host interactions, then following horizontal transmission the recipient host will carry microbes that are closely related to those of the donor host, even when the two hosts are (genetically) unrelated. From the assortment point-of-view, infection by behavior-determining microbes during interactions effectively generates assortment because a recipient of help may be infected by a behavior-determining microbe and consequently become a helper. Cultural

372 horizontal transmission can similarly generate assortment between cooperators and enhance benefit of cooperation if cultural transmission and helping interactions occur between the same individuals, i.e.
374 when there is social association; the recipient of help may also be the recipient of the cultural trait for cooperation. Thus, with horizontal transmission, "assortment between focal cooperative players and
376 cooperative acts in their interaction environment" (**?**) is generated not because *the helper is likely to be helped*, but rather because *the helped is likely to become a helper*.

Acknowledgements

Appendices

Appendix A   Local stability criterion

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

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

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

Therefore, by definition of big-O notation, if $f'(0) < 0$ then there exists $\epsilon > 0$ such that for any local
388 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

$$390 \quad 1 - p' < 1 - p \Leftrightarrow -\frac{f(p)}{1-p} < 0 \Leftrightarrow \frac{f'(1)(p-1) + O\big((p-1)^2\big)}{p-1} < 0 \Leftrightarrow f'(1) - O(1-p) < 0 \ . \tag{A2}$$

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

Appendix B   Equilibria and stability

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

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

where

$$\begin{aligned} \beta_1 &= \big[c(1-v) - b(1-\alpha v)\big](T_A - T_B) \ , \\ 398 \qquad \beta_2 &= -\beta_1 - \beta_3 \ , \\ \beta_3 &= \alpha b v T_A - c v(1 - T_B) + (T_A - T_B) \ . \end{aligned} \tag{B2}$$

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
400 polynomial,

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

11

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

$$
\begin{aligned}
f'(0) &= \alpha bvT - cv(1 - T) \,, \\
f'(1) &= -\alpha bvT + cv(1 - T) \,.
\end{aligned}
\tag{B4}
$$

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

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

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

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

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

Similarly, if $T_A < T_B$, then

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

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

This analysis can be summarized with the following conditions:

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

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

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

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

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

12

442 Second, assume $T_A > T_B$. $\beta_3 > 0$ requires $\gamma_1 > c$. For $\beta_1 < \beta_3$ we need $c\left[v(1-T_B)+(1-v)(T_A-T_B)\right] < bv\alpha T_B+(1+b)(T_A-T_B)$. Thus for $\beta_1 < \beta_3$ we need $v > \hat{v}$ and $c < \gamma_2$ or $v < \hat{v}$ and $c > \gamma_2$. But $\hat{v} < 0$
444 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 **??**.

Appendix C   Effect of social association on mean fitness

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

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

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

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

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

Appendix D   Reduction principle

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

$$
\begin{aligned}
\bar{w}\tilde{p}_1' =\ & \tilde{p}_1 x(1 + b - c)(1 - (1 - \alpha_1)(1 - x)T_B) + \\
& \tilde{p}_1(1 - x)(1 - c)(1 - \alpha_1 T_B x - T_B(1 - x)) + \\
& \tilde{p}_2 x(1 + b)T_A(x + \alpha_1(1 - x)) + \\
& \tilde{p}_2(1 - x)x(1 - \alpha_1)T_A \ , \\
\bar{w}\tilde{p}_2' =\ & \tilde{p}_1 x(1 + b - c)(1 - \alpha_1)(1 - x)T_B + \\
& \tilde{p}_1(1 - x)(1 - c)(\alpha_1 T_B + (1 - \alpha_1)(1 - x)T_B) + \\
& \tilde{p}_2 x(1 + b)(1 - \alpha_1 T_A(1 - x) - T_A x) + \\
& \tilde{p}_2(1 - x)(1 - (1 - \alpha_1)xT_A) \ , \\
\bar{w}\tilde{p}_3' =\ & \tilde{p}_3 x(1 + b - c)(1 - (1 - \alpha_2)(1 - x)T_B) + \\
& \tilde{p}_3(1 - x)(1 - c)(1 - \alpha_2 T_B x - T_B(1 - x)) + \\
& \tilde{p}_4 x(1 + b)T_A(x + \alpha_2(1 - x)) + \\
& \tilde{p}_4(1 - x)x(1 - \alpha_2)T_A \ , \\
\bar{w}\tilde{p}_4' =\ & \tilde{p}_3 x(1 + b - c)(1 - \alpha_2)(1 - x)T_B + \\
& \tilde{p}_3(1 - x)(1 - c)(\alpha_2 T_B + (1 - \alpha_2)(1 - x)T_B) + \\
& \tilde{p}_4 x(1 + b)(1 - \alpha_2 T_A(1 - x) - T_A x) + \\
& \tilde{p}_4(1 - x)(1 - (1 - \alpha_2)xT_A) \ ,
\end{aligned}
\tag{D1}
$$

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

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

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

464 which we get by setting $\alpha = \alpha_1$ and $v = 1$ in **??**. When $v = 1$, this equilibrium is polymorphic ($0 < \tilde{p}^* < 1$) if $T_A < T_B$ and $\gamma_2 < c < \gamma_1$ (Result **??**).

466 The local stability of $\tilde{\mathbf{p}}^*$ to the introduction of allele m is determined by the linear approximation $\mathbf{L}^*$ of the transformation in **??** near $\tilde{\mathbf{p}}^*$ (i.e., the Jacobian of the transformation at the equilibrium). $\mathbf{L}^*$ is 468 known to have a block structure, with the diagonal blocks occupied by the matrices $\mathbf{L}^*_{in}$ and $\mathbf{L}^*_{ex}$. The latter is the external stability matrix: the linear approximation to the transformation near $\tilde{\mathbf{p}}^*$ involving 470 only the pheno-genotypes Am and Bm, derived from **??** as

$$\mathbf{L}^*_{ex} = \frac{1}{\bar{w}^*} \begin{bmatrix} X & Y \\ Z & Q \end{bmatrix} = \frac{1}{\bar{w}^*} \begin{bmatrix} \frac{\partial \bar{w}\tilde{p}'_3}{\partial \tilde{p}_3}(\tilde{\mathbf{p}}^*) & \frac{\partial \bar{w}\tilde{p}'_3}{\partial \tilde{p}_4}(\tilde{\mathbf{p}}^*) \\ \frac{\partial \bar{w}\tilde{p}'_4}{\partial \tilde{p}_3}(\tilde{\mathbf{p}}^*) & \frac{\partial \bar{w}\tilde{p}'_4}{\partial \tilde{p}_4}(\tilde{\mathbf{p}}^*) \end{bmatrix} =$$

$$\frac{1}{\bar{w}^*} \begin{bmatrix} (1 + b\tilde{p}^* - c)(1 - T_B(1 - \tilde{p}^*)) + b\tilde{p}^*\alpha_2 T_B(1 - \tilde{p}^*) & (1 + b\tilde{p}^*)T_A\tilde{p}^* + b\tilde{p}^*\alpha_2 T_A(1 - \tilde{p}^*) \\ (1 + b\tilde{p}^* - c)T_B(1 - \tilde{p}^*) - b\tilde{p}^*\alpha_2 T_B(1 - \tilde{p}^*) & (1 + b\tilde{p}^*)(1 - T_A\tilde{p}^*) - b\tilde{p}^*\alpha_2 T_A(1 - \tilde{p}^*) \end{bmatrix}. \tag{D3}$$

472 Because we assume that $\tilde{\mathbf{p}}^*$ is internally stable (i.e. locally stable to small perturbations in the frequencies of AM and BM), the stability of $\tilde{\mathbf{p}}^*$ is determined by the eigenvalues of the external 474 stability matrix $\mathbf{L}^*_{ex}$. This is a positive matrix, and due to the Perron-Frobenius theorem, the leading eigenvalue of $\mathbf{L}^*_{ex}$ is real and positive. Thus, if the leading eigenvalue is less (greater) than one, then 476 the equilibrium $\tilde{\mathbf{p}}^*$ is externally stable (unstable) and allele m cannot (can) invade the population of allele M. The eigenvalues of $\mathbf{L}^*_{ex}$ are the roots of the characteristic polynomial,

$$R(\lambda) = \lambda^2 - \lambda \frac{(X + Q)}{\bar{w}^*} + \frac{XQ - YZ}{\bar{w}^{*2}}, \tag{D4}$$

where X, Y, Z, and Q are defined in **??**. The characteristic polynomial $R(\lambda)$ is a quadratic with a 480 positive leading coefficient. Therefore, $\lim_{\lambda \to \pm\infty} R(\lambda) = \infty$, and the leading eigenvalue is less than one (implying stability) if and only if $R(1) > 0$ and $R'(1) > 0$. Thus, a sufficient condition for external 482 instability of $\tilde{\mathbf{p}}^*$ is

$$R(1) < 0 \Leftrightarrow 1 - \frac{(X + Q)}{\bar{w}^*} + \frac{XQ - YZ}{\bar{w}^{*2}} < 0 \Leftrightarrow \bar{w}^{*2} - \bar{w}^*(X + Q) + XQ - YZ < 0. \tag{D5}$$

484 Using *SymPy* (**?**), a Python library for symbolic mathematics, we find that inequality **??** is true if and only if

486 $$\alpha_2\big(bT_A + (T_A - T_B) - c(1 - T_B)\big) < \alpha_1\big(bT_A + (T_A - T_B) - c(1 - T_B)\big). \tag{D6}$$

We assumed $c < \gamma_1$, that is,

488 $$\begin{aligned} c < \gamma_1 &= \frac{b\alpha_1 T_A + (T_A - T_B)}{1 - T_B} \Leftrightarrow \\ 0 &< b\alpha_1 T_A + (T_A - T_B) - c(1 - T_B) \Rightarrow \\ 0 &< bT_A + (T_A - T_B) - c(1 - T_B), \end{aligned} \tag{D7}$$

since $0 \le \alpha_1 \le 1$. Therefore, combining inequalities **??**, **??**, and **??**, we find that $R(1) < 0$ if and only 490 if $\alpha_2 < \alpha_1$. This is a sufficient condition for external instability, so, if the social association α_2 of the invading modifier allele m is less than α_1 of the resident M, then invasion will be successful.

492 Determining a necessary and sufficient condition for successful invasion is more complicated, requiring analysis of the sign of $R'(1)$. However, we validate using numerical simultations that $R(1) > 0$ when 494 $\alpha_2 < \alpha_1$ for different values of b, TA, TB, c and $\alpha_1 > \alpha_2$. In addition, we also validate that $R(1) < 0$ when $\alpha_2 > \alpha_1$ for different values of b, TA, TB, c and $\alpha_1 < \alpha_2$.

14

Tables

Table 1: Payoff matrix for prisoner's dilemma.

	$\phi_2 = A$	$\phi_2 = B$
$\phi_1 = A$	$1 + b - c$	$1 - c$
$\phi_1 = B$	$1 + b$	1

The fitness of phenotype ϕ_1 when interacting with phenotype ϕ_2. A is a cooperative phenotype, B is a defector phenotype, b is the benefit gained by an individual interacting with a cooperator, and c is the cost of cooperation. $0 < c < 1$ and $c < b$.

Table 2: 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	$P(\phi_1 = A)$ via horizontal transmission: 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 3: Model variables and parameters.

| Symbol	Description	Values
A	Cooperator phenotype	
B	Defector phenotype	
p	Frequency of phenotype A among adults	$[0, 1]$
$\tilde{p}$	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 social association	$[0, 1]$
T_A, T_B	Horizontal transmission rates of phenotype A and B	$[0, 1]$

Figures

Supplementary Material
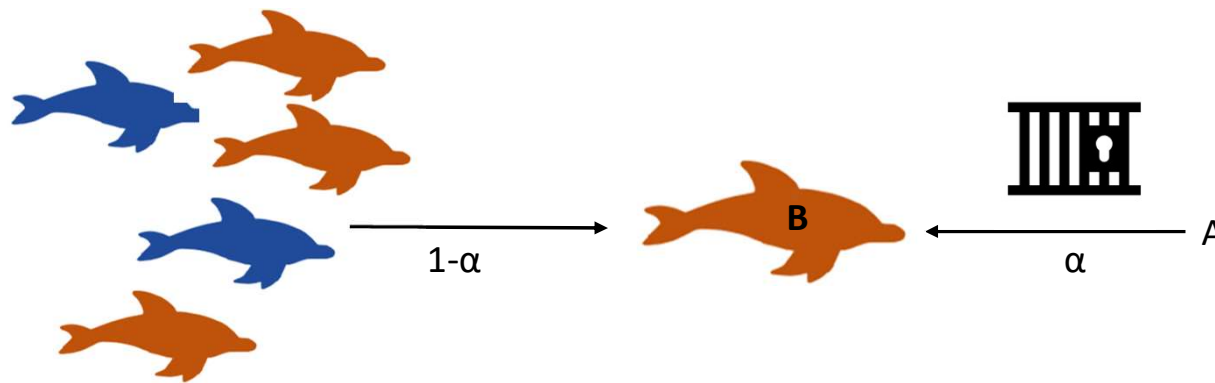
Figure 1: Cultural horizontal transmission with assortment. Transmission occurs between inter-
acting partners with probability α (left) or between two random peers with probability $1 - \alpha$, where
α is the *social association* parameter.
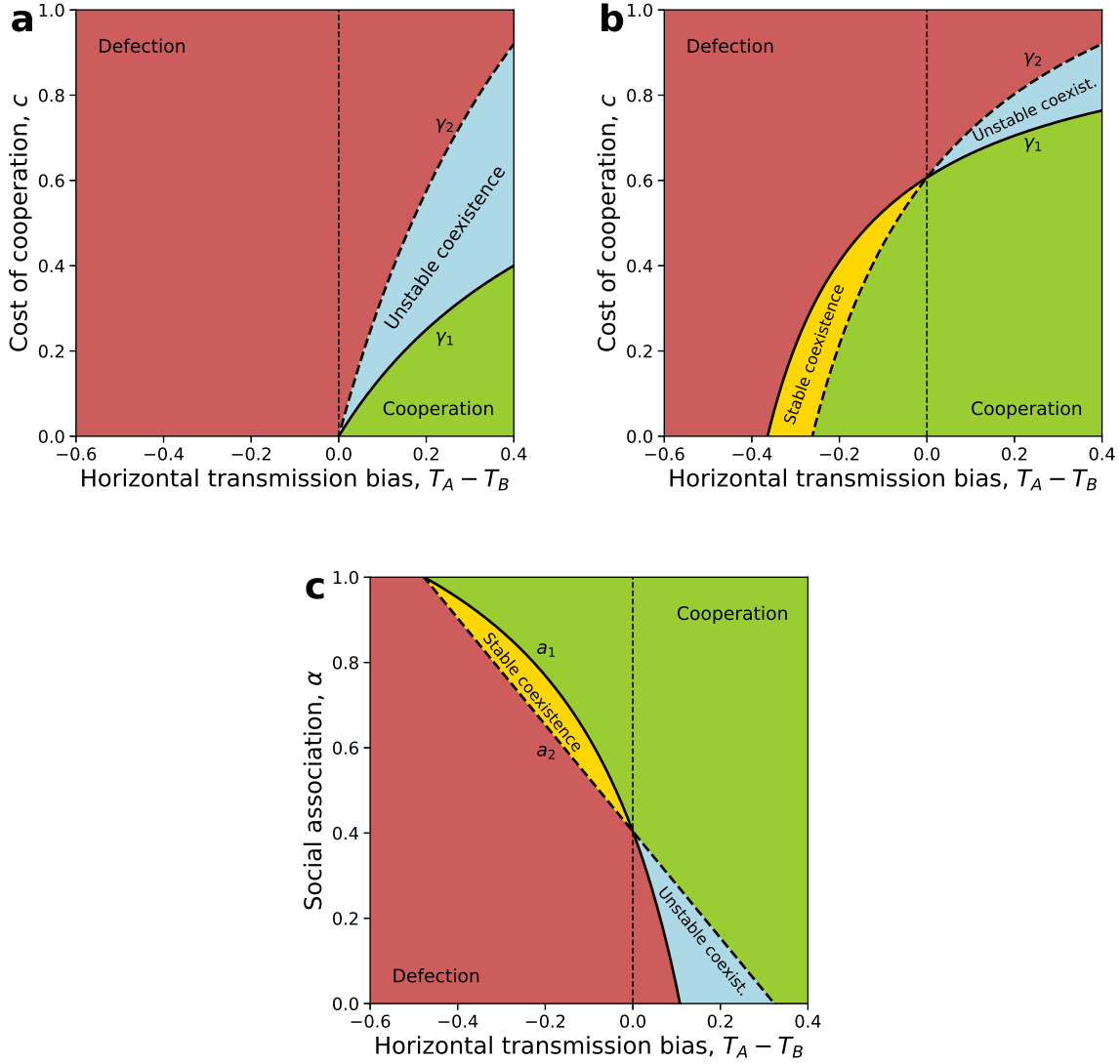
Figure 2: Evolution of cooperation under vertical and horizontal cultural transmission. The figure shows the parameter ranges for global fixation of cooperation (green), global fixation of defection (red), fixation of either cooperation or defection depending on the initial conditions, i.e. unstable polymorphism (blue), and stable polymorphism of cooperation and defection (yellow). In all cases the horizontal bias ($T_A - T_B$) is on the x-axis. (**a-b**) Cost of cooperation c is on the y-axis; the cost thresholds γ_1 and γ_2 (**??**) are the solid and dashed lines, respectively. (**c**) Social association α is on the y-axis; the social association thresholds a_1 and a_2 (**??**) are the solid and dashed lines, respectively. Here, $b = 1.3$, $T_A = 0.4$, (**a**) $\alpha = 0$, (**b**) $\alpha = 0.7$, (**c**) $c = 0.35$.
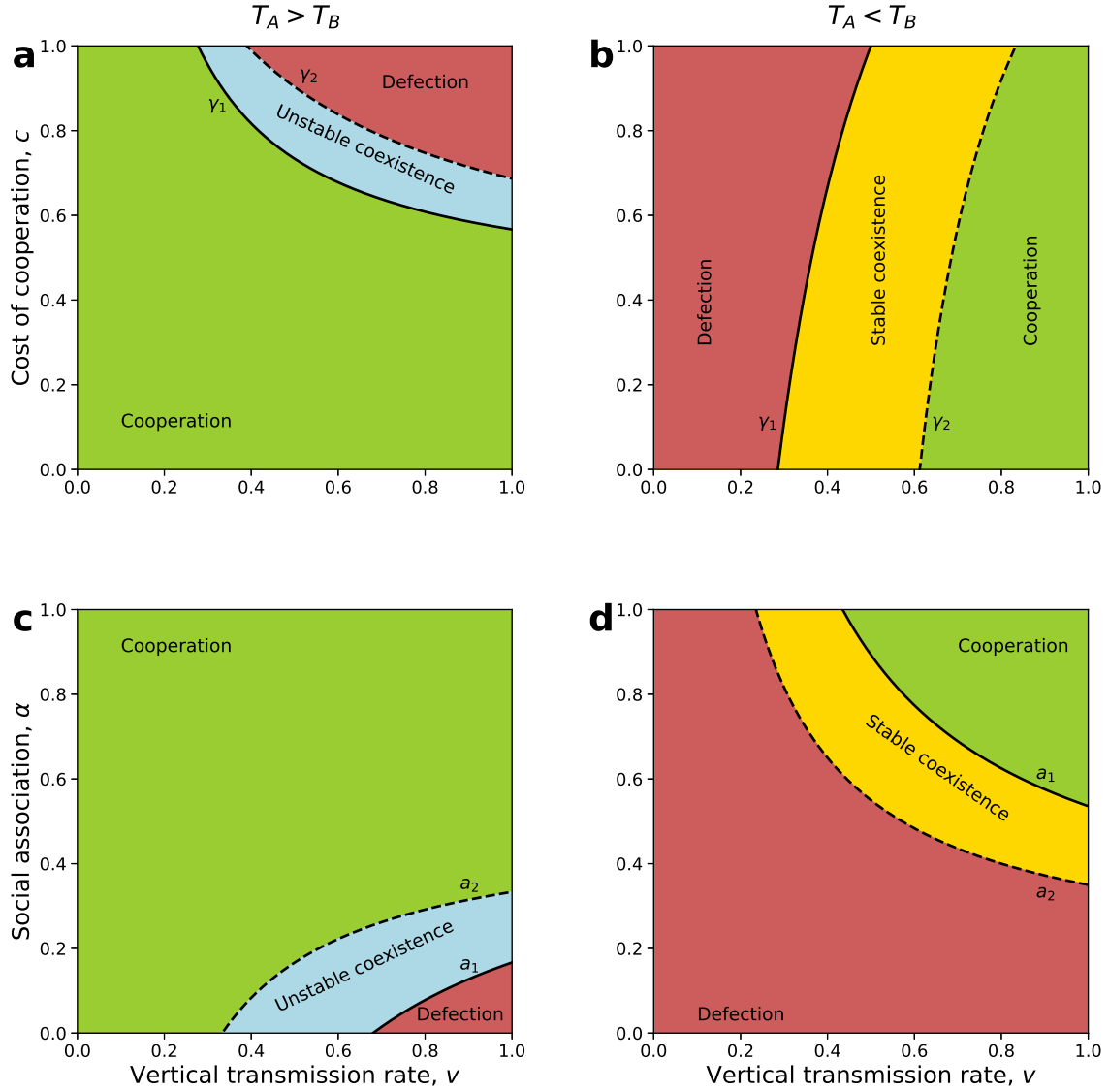
Figure 3: Evolution of cooperation under vertical, oblique, and horizontal cultural transmission.
The figure shows the parameter ranges for global fixation of cooperation (green), global fixation of defection (red), fixation of either cooperation or defection depending on the initial conditions, i.e. 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 (**??**) are represented by the solid and dashed lines, respectively. (**c-d**) Social association α is on the y-axis and the social association thresholds a_1 and a_2 (**??**) are represented by the solid and dashed lines, respectively. Horizontal transmission is biased in favor of cooperation, $T_A > T_B$, in (**a**) and (**c**), or defection, $T_A < T_B$, in (**b**) and (**d**). Here, $T_A = 0.5$, and (**a**) $b = 1.2$, , $T_B = 0.4$, $\alpha = 0.4$; (**b**) $b = 2$, $T_B = 0.7$, $\alpha = 0.7$; (**c**) $b = 1.2$, $T_B = 0.4$, $c = 0.5$; (**d**) $b = 2$, $T_B = 0.7$, $c = 0.5$.
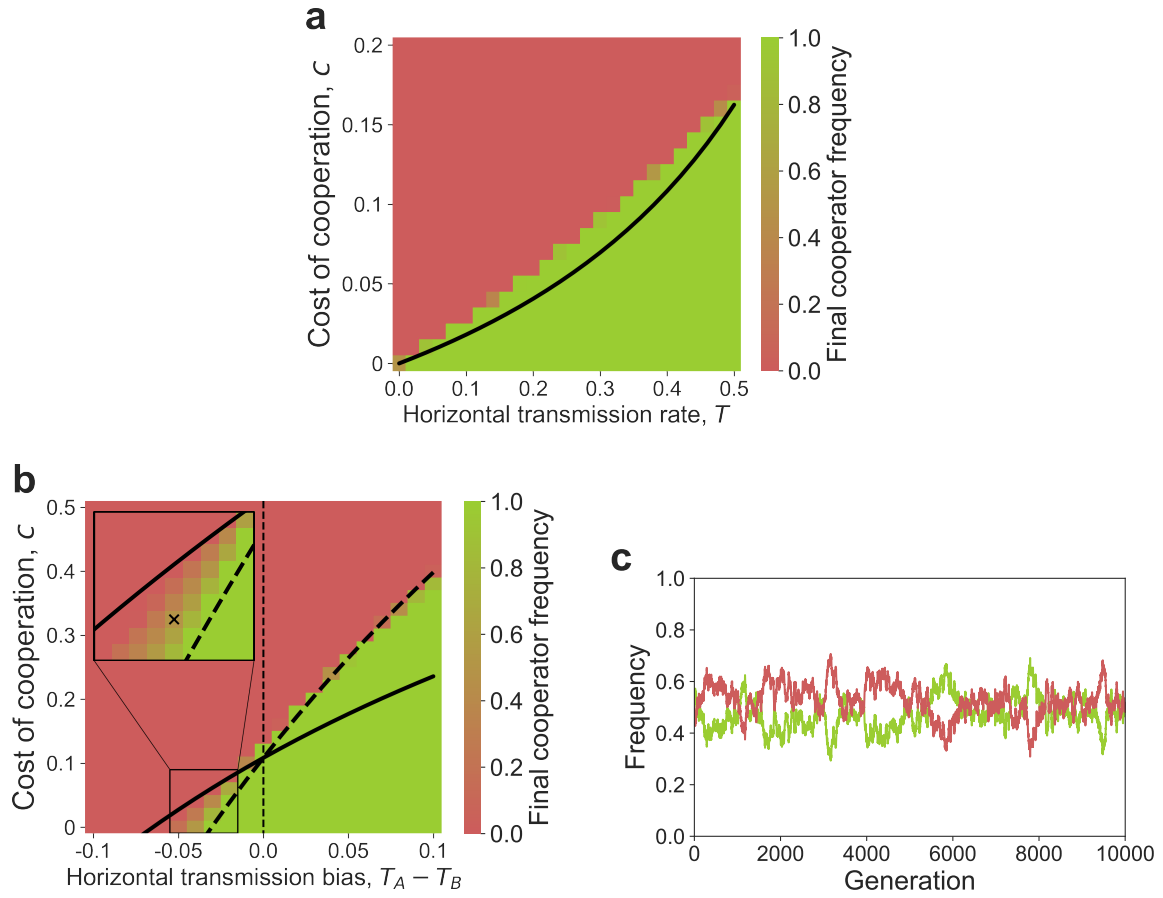
Figure 4: Evolution of cooperation in a structured population. (a-b) The expected frequency of cooperators in a structured population after 10,000 generations is shown (red for 0%, green for 100%) as a function of both the cost of cooperation, c, on the y-axis, and either the symmetric horizontal transmission 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 social association where $\alpha = 1/8$ in **??** for panel (a) and in **??** for (b). **(b)** The inset focuses on an area of the parameter range in which neither phenotype is fixed throughout the simulation, maintaining a stochastic locally stable polymorphism (**?**). **(c)** 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 neighboring sites, and each site has 8 neighbors. Selection operates globally (see **??** 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 rates $T_A = 0.4$, $0.3 < T_B < 0.5$; **(c)** Horizontal transmission rates $T_A = 0.4 < T_B = 0.435$ and cost of cooperation $c = 0.02$.
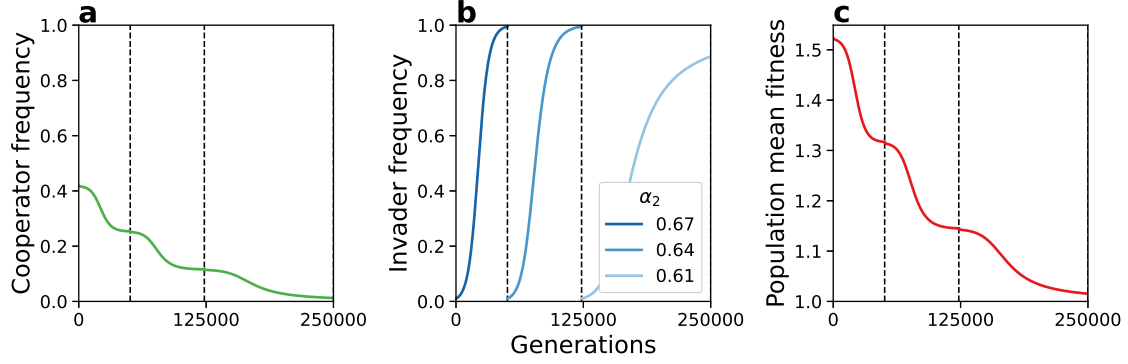
Figure 5: Reduction principle for social association. Consecutive fixation of modifier alleles that reduce social association in numerical simulations of evolution with two modifier alleles (**??**). When a modifier allele fixes (frequency > 99.95%), a new modifier allele that reduces social association by 5% is introduced. **(a)** The frequency of the cooperative phenotype A in the population over time. **(b)** The frequency of the invading modifier allele m over time. **(c)** The population mean fitness over time. Invading alleles are introduced at frequency 0.5%. Here, $c = 0.05$, $b = 1.3$, $T_A = 0.4 < T_B = 0.7$, initial social association $\alpha_1 = 0.7$.
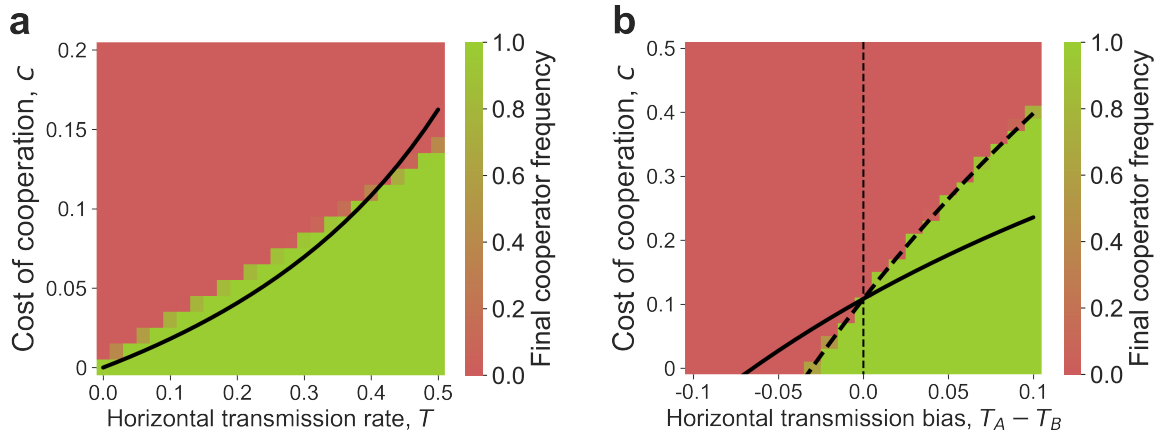
Figure S1: Evolution of cooperation in a structured population with local selection. The expected frequency of cooperators in a structured population after 10,000 generations is shown (red for 0%, green for 100%) as function of both the cost of cooperation (c) on the y-axis, and the symmetric horizontal transmission rate ($T = T_A = T_B$) on the x-axis of the left panel, or the transmission bias $T_A - T_B$ on the x-axis of the right panel. The population evolves on a 100-by-100 grid. Cooperation and horizontal transmission are both local between neighboring sites, and each site had 8 neighbors. Selection operates locally (see **??** 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 social association where $\alpha = 1/8$ in **(a) ??** and **(b) ??**. Simulations were stopped at generation 10,000 or if one of the phenotypes fixed and 50 simulations were executed for each parameter set. Here, population size is 10,000 (100-by-100 grid), benefit of cooperation, $b = 1.3$, perfect vertical transmission $v = 1$. **(a)** Symmetric horizontal transmission, $T = T_A = T_B$. **(b)** Horizontal transmission rates $T_A = 0.4$, $0.3 < T_B < 0.5$.
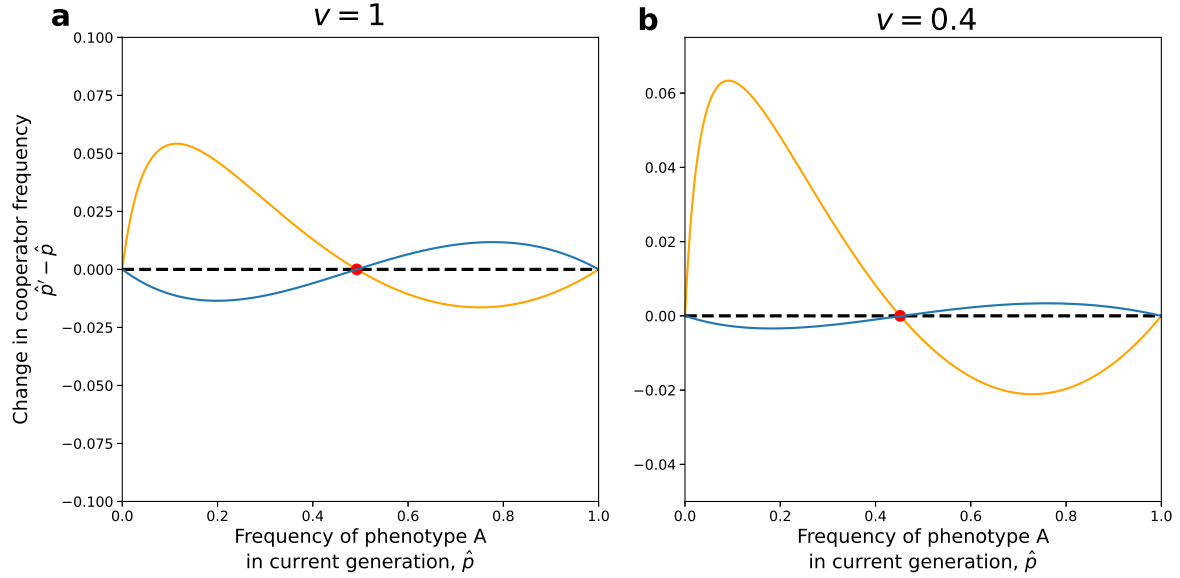
Figure S2: Stable and unstable polymorphism of cooperation and defection. The curves show the difference of $\hat{p}' - \hat{p}$ vs $\hat{p}$ where $\hat{p}'$ abd $\hat{p}$ are the frequency of the cooperative phenotype A among juveniles in the next generation and current generations (**??**). The dashed black line is $\hat{p}' - \hat{p} = 0$. The curves and the dashed line intersect at the stable equilibrium $\hat{p}^*$ (red circle). When $\hat{p}' - \hat{p} > 0$, $\hat{p}$ increases towards $\hat{p}^*$. When $\hat{p}' - \hat{p} < 0$, $\hat{p}$ decreases towards $\hat{p}^*$. **(a)** There is no oblique transmission, $v = 1$. The orange curve, for which the polymorphic equilibrium is stable, is given by $T_A = 0.4$, $T_B = 0.9$, $b = 12$, $c = 0.35$, and $\alpha = 0.45$, which give $\gamma_2 < c < \gamma_1$ (**??**). The blue curve, for which the equilibrium is unstable, is given by $T_A = 0.5$, $T_B = 0.1$, $b = 1.3$, $c = 0.904$, and $\alpha = 0.4$, which give $\gamma_1 < c < \gamma_2$. **(b)** Oblique transmission exists. The orange curve is parameterized by $T_A = 0.4$, $T_B = 0.9$, $b = 20$, $c = 0.1$, $\alpha = 1$, and $v = 0.4$, which give $0 < \beta_3 < \beta_1$ (**??**). The blue curve is parameterized by $T_A = 0.5$, $T_B = 0.4$, $b = 1.2$, $c = 0.487$, $\alpha = 0.09$ and $v = 0.6$, which give $\beta_1 < \beta_3 < 0$.
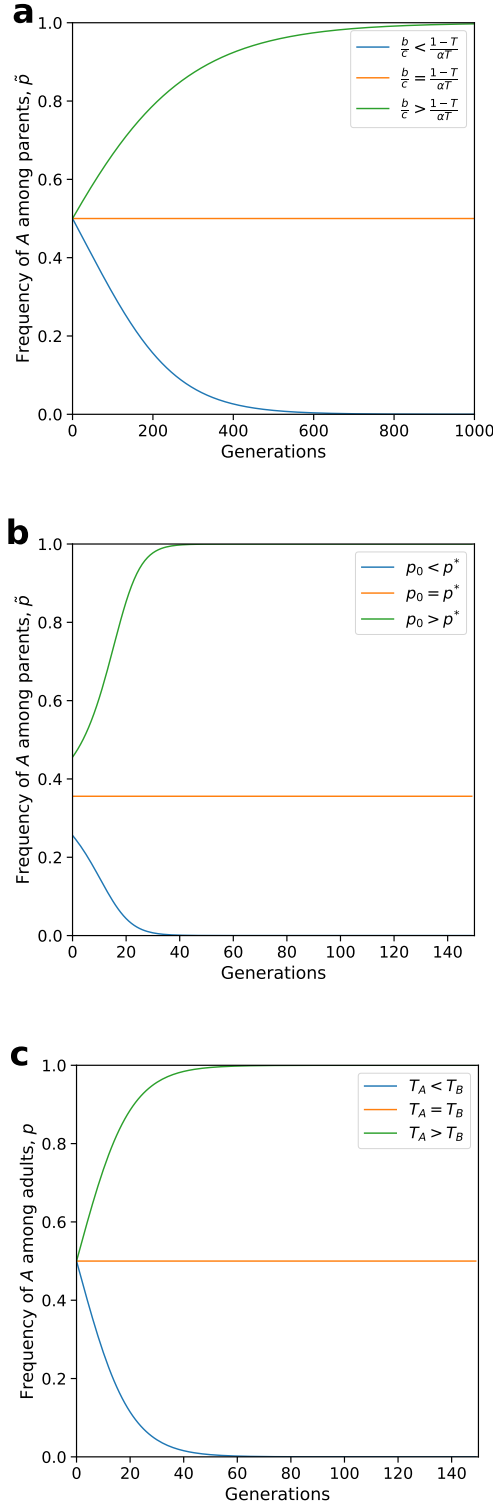
Figure S3: Dynamics of the frequency of cooperation. The frequency $\tilde{p}$ of cooperator parents in **(a-b)** and the frequency p of adults cooperators in **(c)**. The different lines correspond to parameter values that lead to fixation of cooperation (green), extinction of cooperation (red), or stable polymorphism of cooperators and defectors (yellow). **(a)** $v = 1$, $T_A = T_B = T = 0.2$, $\alpha = 0.5 \neq 0$, $\tilde{p}_0 = 0.5$ and $c = 0.1$; **(b)** $v = 1$, $\alpha = 0$, $\tilde{p}^* \approx 0.35$, $T_A = 0.65$, $T_B = 0.1$, $b = 1.3$ and $c = 0.65$; **(c)** $v = 0$, $\alpha = 0.5$, $p_0 = 0.5$, $T_A = 0.5$, $b = 1.3$ and $c = 0.5$.