WORKING PAPERS

N° 1737

April 2026

“Evolutionary branching of social preferences in a public good provision game”

Guillaume Cheikbossian and Jorge Peña

Toulouse School of Economics

Evolutionary branching of social preferences in a public good provision game

Guillaume Cheikbossian*[1] and Jorge Peña[2]

[1]CEE-M, Université de Montpellier, CNRS, INRAE, Institut Agro, Montpellier, France
[2]Toulouse School of Economics, Universite Toulouse Capitole, Toulouse, France

Abstract

We study the evolution of other-regarding preferences in a public goods game where the production function exhibits varying degrees of complementarity between individual efforts. Individuals are rational agents who play a Nash equilibrium, but differ in the weight they assign to others' payoffs, capturing varying degrees of prosocial or anti-social preferences. This preference trait evolves through payoff-based biased social learning, modeled within an adaptive dynamics framework. Because material payoffs induced by the equilibrium contributions may be non-concave in the preference parameter, evolutionary branching can arise. We show that monomorphic populations are evolutionarily stable only when complementarity between individual efforts is sufficiently strong, in which case preferences converge toward either prosociality or anti-sociality depending on the nature of strategic interactions between players. By contrast, when contributions are highly substitutable, monomorphic populations can become unstable, giving rise to polymorphic populations in which multiple preference types coexist. These results highlight how the structure of the public goods environment shapes the evolution and diversity of other-regarding motivations in culturally evolving populations.

Keywords: adaptive dynamics; other-regarding preferences; public goods games.

---

*Corresponding author: guillaume.cheikbossian@umontpellier.fr

1 Introduction

Individuals in social species differ persistently in their cooperative behavior. In nature, populations of microorganisms and animals routinely exhibit stable coexistence of individuals that invest heavily in collective goods alongside others that contribute little or nothing (Griffin et al., 2004; Gore et al., 2009; Doebeli et al., 2004; Archetti and Scheuring, 2011). In humans, experiments on public goods provision consistently reveal substantial and repeatable heterogeneity in contributions, even among participants facing identical incentives (Chaudhuri, 2011; Fehr and Schmidt, 1999). A central question in evolutionary theory is whether such behavioral diversity can be understood as an adaptive outcome: can evolution, acting on a population that is initially homogeneous, generate and maintain multiple coexisting types that differ in their propensity to cooperate?

The framework of *adaptive dynamics* (Dieckmann and Law, 1996; Geritz et al., 1998) provides the tools to address this question for continuously varying traits. A key insight of adaptive dynamics is the distinction between two stability properties of evolutionary equilibria. A trait value may be *convergence stable*, meaning that gradual evolution under recurrent mutation drives the population toward it, yet *evolutionarily unstable*, meaning that once reached, such trait value can be invaded by nearby mutants on either side. When both conditions hold simultaneously, the population undergoes *evolutionary branching*: an initially monomorphic population splits into distinct coexisting phenotypic clusters (Geritz et al., 1998). Applying this framework to strategy evolution in public goods games, Doebeli et al. (2004) showed that populations can diversify from uniform intermediate investment levels into coexisting cooperators and defectors. More recently, Lehmann and Mullon (2025) developed a general analysis of adaptive polymorphisms in quantitative traits, identifying *negative trait-dependent selection*—whereby population-level expression of a trait weakens selection on that trait—as the unifying condition for the gradual emergence of polymorphism across ecological, social, and genetic interactions.

Most applications of adaptive dynamics in biology study the evolution of strategies or behaviors directly. However, behavior in social interactions often arises as the equilibrium outcome of a process in which individuals optimize according to their underlying motivations or preferences (McNamara et al., 1999; Akçay et al., 2009; Alger et al., 2020; Alger, 2023). In such cases, it is not the behavior itself but the preference that is the heritable trait under selection. This perspective is formalized in the *indirect evolutionary approach* (Güth and Yaari, 1992; Bester and Güth, 1998), developed mainly in economics: evolution operates not on strategies but on the parameters of individuals' utility functions, while behavior is determined by rational optimization given those preferences. An important class of preferences studied in this literature is *other-regarding preferences*, where individuals place positive or negative weight on the material payoffs of others (Bester and Güth, 1998; Sethi and Somnathan, 2001; Ok and Vega-Redondo, 2001; Heiftez et al., 2007b,a; Dekel et al., 2007; Alger and Weibull, 2013, 2019). This weight—capturing varying degrees of benevolence or malevolence—can be interpreted as a genetically or culturally inherited disposition shaping cooperative behavior, analogous to the response rules and social motivations studied in behavioral ecology (McNamara et al., 1999; Akçay and Cleve, 2012).

Previous analyses of preference evolution have focused on identifying evolutionarily stable preferences using the static ESS concept (Alger, 2023). These analyses face two limitations. First, they typically assume, often implicitly, that the fitness function is concave in the preference parameters, which guarantees the existence of a fitness-maximizing preference type. Yet concavity of utility in actions does not imply concavity of the induced material payoff—and hence fitness—in the preference parameter. There is no general reason why fitness-maximizing preferences should exist even in simple strategic environments. Second, the ESS concept says nothing about convergence stability: it identifies trait values that resist invasion but does not describe whether evolution actually reaches them, nor whether disruptive selection can split a population into coexisting types. These two limitations are, in fact, related. When fitness is non-concave in the preference trait, the singular point of the adaptive dynamics can be convergence stable yet evolutionarily unstable, which is precisely the condition for evolutionary branching.

The contribution of this paper is to apply the full machinery of adaptive dynamics to the evolution of other-regarding preferences in a public goods game. We model the joint production of the public good using a *constant elasticity of substitution (CES)* aggregator of individual efforts (Cornes and Hartley, 2007; Buchholz and Sandler, 2021), which allows us to vary the degree of complementarity or substitutability among contributions continuously. The evolving trait is the weight that individuals place on others' material payoffs. While individual utility is concave in effort—guaranteeing a unique Nash equilibrium in contributions—the induced material payoffs can be non-concave in the preference parameter, depending on the degree of complementarity and group size. Our key finding is that a monomorphic population is evolutionarily stable only when complementarity between efforts is sufficiently strong. In this case, preferences evolve toward benevolence or malevolence depending on the nature of strategic interactions. When contributions are highly substitutable, however, the singular preference trait becomes an evolutionary branching point, and the population diversifies into coexisting types with heterogeneous social preferences. Larger groups dampen this tendency toward polymorphism, favoring convergence to a homogeneous population with preferences closer to selfishness. Importantly, evolutionary branching occurs at the preference level in a setting where it *cannot* occur at the strategy level, since material payoffs are strictly concave in effort. This highlights a mechanism for the evolutionary origin of behavioral diversity that operates through the non-concavity of fitness in preferences rather than through the structure of the payoff function in strategies, as in Doebeli et al. (2004).

The remainder of the paper is organized as follows. Section 2 presents the public goods game with other-regarding preferences and characterizes the Nash equilibrium in contributions. Section 3 applies adaptive dynamics to the preference trait and identifies the conditions for evolutionary stability, convergence stability, and branching, supplemented by individual-based simulations. Section 4 discusses our results in relation to strategy-level branching (Doebeli et al., 2004) and to the recent analysis of preference evolution in linear-quadratic games by Lehmann and Mullon (2025). Section 5 concludes.

2 Public good production

We consider a very large population of players. In each period, $n \geq 2$ players are matched uniformly at random for the joint production of a certain output having the characteristics of a public good. Unlike most studies of public goods games (which assume that individual contributions are perfectly substitutable) we model the public-good production as a CES function of individual efforts with a variable degree of complementarity.[1] We assume that all players have the same finite endowment $y \in \mathbb{R}_+$. Let $\mathbf{e} = (e_1, \ldots, e_n)$ denote the vector of individual efforts, where $e_i \in [0, y]$ for $i = 1, \ldots, n$ is the effort of the ith player. The production function of the public good, $G : [0, y]^n \to \mathbb{R}_+$, is given by

$$G(\mathbf{e}) = \left(\frac{1}{n} \sum_{j=1}^{n} e_j^{1-\sigma} \right)^{\frac{1}{1-\sigma}}, \tag{1}$$

where $\sigma \in [0, 1) \cup (1, \infty)$ is a measure of the degree of complementarity between individual efforts.[2] The elasticity of substitution is $1/\sigma$ so that the higher σ, the higher is the degree of complementarity.[3] For $\sigma = 0$, we have perfect substitutability between individual efforts,

---

[1] Following the canonical model of Bergstrom et al. (1986), most studies of non-cooperative public goods games assume that the aggregate level of public good provision is the unweighted sum of all individuals' contributions. Cornes and Hartley (2007) were the first to study public goods games with a varying degree of complementarity between individual efforts, using the CES aggregator and focusing on equilibrium existence and uniqueness. Nevertheless, this type of model remains largely unexplored, particularly for technical reasons. For a recent survey with a special emphasis on aggregator technologies in public goods games, see Buchholz and Sandler (2021).

[2] An alternative CES production function is $G(\mathbf{e}) = \left(\sum_{j=1}^{n} e_j^{1-\sigma} \right)^{\frac{1}{1-\sigma}}$ (Adams, 2006; Ray et al., 2007; Cheikbossian and Fayat, 2018). In this case, with a common level of effort e, we have $G(\mathbf{e}) = n^{\frac{1}{1-\sigma}} e$ and total production mechanically increases (resp. decreases) with group size n when $\sigma \in [0, 1)$ (resp. $\sigma \in (1, \infty)$). We prefer to use Eq. (1) to isolate the degree of complementarity of individual efforts from any additional effect of positive or negative synergy induced by an increase in group size according to the value of σ. However, even if this alternative CES production function changes the levels of effort at the Nash equilibrium given the social preferences of the agents, it does not change any of the results relating to the convergence and stability of the evolution of preferences.

[3] In the economics literature, the elasticity of substitution is defined as the percentage change in the ratio of two inputs used, divided by the percentage change in their *marginal rate of technical substitution* (MRTS). The MRTS is defined as the rate at which an input can be substituted for another input while holding total output constant and is, thus, given by the ratio of marginal productivities of inputs. Thus, the elasticity of substitution between two inputs e_i and e_j is formally given by $\epsilon_{ij} = \frac{d(e_i/e_j)/(e_i/e_j)}{d(G_i/G_j)/(G_i/G_j)}$, where G_i (resp. G_j) is

and Eq. (1) becomes the standard "summation technology", i.e., $G(\mathbf{e}) = (1/n)\sum_{j=1}^{n} e_j$. For $\sigma \to +\infty$, we have perfect complementarity between individual efforts, and Eq. (1) becomes the "weakest-link technology", i.e., $G(\mathbf{e}) = \min\{e_1, \ldots, e_n\}$. Finally, if $\sigma > 1$ and there exists j for whom $e_j = 0$, and the production function $G(\mathbf{e})$ is not well-defined. Hence, we will take the limit of $G(\mathbf{e})$ as $e_j \to 0$ in this case, which is $G(\mathbf{e}) = 0$.[4]

Player i's *material payoff* is

$$\Pi_i(\mathbf{e}) = [G(\mathbf{e})]^{\alpha} - c(e_i), \tag{2}$$

where $\alpha \in (0, 1)$ is a curvature parameter governing concavity in public-good consumption. We thus assume, as usual, that there are decreasing returns to scale in satisfaction obtained by consuming the public good. The cost of exerting effort e_i is given by $c(e_i)$, where $c : [0, y] \to \mathbb{R}_+$ is the cost function which, as usual again, we assume to be convex so that the marginal cost of effort is increasing. To simplify the analysis, we focus on the cost function

$$c(e_i) = (1/\beta)e_i^{\beta}, \tag{3}$$

where $\beta \in [1, \infty)$ is the (constant) elasticity of effort cost.

Players have other-regarding preferences, with *utility functions* depending on both their own material payoff and the material payoffs of co-players. The utility of player i is given by

$$V_i(\mathbf{e}) = \Pi_i(\mathbf{e}) + \lambda_i \sum_{\substack{j=1 \\ j \neq i}}^{n} \Pi_j(\mathbf{e}), \tag{4}$$

where $\lambda_i \in (-1/(n-1), 1]$ is the utility weight given by player i to the material payoffs of co-players. Positive values represent "benevolence", negative values represent "malevolence", and $\lambda_i = 0$ corresponds to "egoism".[5]

Substituting Eq. (2) and Eq. (3) into Eq. (4), we have

$$V_i(\mathbf{e}) = \theta_i[G(\mathbf{e})]^{\alpha} - \frac{1}{\beta}e_i^{\beta} - \frac{\lambda_i}{\beta}\sum_{\substack{j=1 \\ j \neq i}}^{n} e_j^{\beta}, \tag{5}$$

where

$$\theta_i \equiv 1 + \lambda_i(n-1) \in (0, n]. \tag{6}$$

---

the derivative of the production function with respect to e_i (resp. e_j). Using Eq. (1), we have that $\epsilon_{ij} = 1/\sigma$, $\forall (e_i, e_j) \in [0, y]^2$.

[4] For $\sigma \to 1$, Eq. (1) converges to the Cobb-Douglas aggregator $G(\mathbf{e}) = \left(\prod_{j=1}^{n} e_j\right)^{1/n}$, i.e. to the geometric mean of individual efforts.

[5] In the economics literature, the terms "altruism" and "spitefulness" are more commonly used (e.g. (Bester and Güth, 1998)). To avoid confusion about the meaning of these terms in evolutionary biology, we use "benevolence" and "malevolence" instead. Note also that we assume $\lambda_i > -1/(n-1)$, since otherwise there would be a complete absence of effort in a symmetric equilibrium. At the other extreme, when $\lambda_i = 1$, player i values the material payoffs of his partners as much as his own payoff.

Each player i chooses e_i to maximize $V_i(\mathbf{e})$, given by (5). We have the following result.[6]

Proposition 1. *Define* $\theta_i \equiv 1 + \lambda_i(n-1) > 0$ *for* $i = 1, \ldots, n$, *and* $\mathbf{\Theta} \equiv (\theta_1, \ldots, \theta_n) \in (0, n]^n$. *Let us further assume that* $\beta + \sigma - 1 \neq 0$ *holds. Then, the public good game admits a unique interior Nash equilibrium with strictly positive effort levels,* $e_i^* > 0$, *for* $i = 1, \ldots, n$, *which satisfies*

$$e_i^*(\mathbf{\Theta}) = \alpha^{\frac{1}{\beta-\alpha}} \left(\frac{1}{n}\right)^{\frac{\alpha}{(\beta-\alpha)(1-\sigma)}} \theta_i^{\frac{1}{\beta+\sigma-1}} \left(\sum_{j=1}^{n} \theta_j^{\frac{1-\sigma}{\beta+\sigma-1}}\right)^{\frac{\alpha+\sigma-1}{(1-\sigma)(\beta-\alpha)}}. \tag{7}$$

When $\beta + \sigma - 1 = 0$, implying perfect substitutability ($\sigma = 0$) together with constant marginal cost ($\beta = 1$), there may exist one or more equilibria in which a subset of players is active and exerts positive effort, while the other players free-ride completely. The active players are those with the highest level of social preference among all players. Also, when $\sigma \to +\infty$ (that is, under perfect complementarity), there exists a continuum of equilibria where all players exert the same effort level.[7] From now on, we abstract from these extreme cases of perfect substitutability or complementarity of individual efforts and focus on the unique interior equilibrium identified in Proposition 1.

One can observe that the equilibrium level of effort of each player, given by Eq. (7), is increasing in his preference parameter θ_i and that it is decreasing in group size. Now, substituting the equilibrium effort values into Eq. (2), and simplifying, we obtain the following expression for player i's equilibrium material payoff:

$$\Pi_i^*(\mathbf{\Theta}) = \alpha^{\frac{\alpha}{\beta-\alpha}} \left(\frac{1}{n}\right)^{\frac{\alpha\beta}{(\beta-\alpha)(1-\sigma)}} \left(\sum_{j=1}^{n} \theta_j^{\frac{1-\sigma}{\beta+\sigma-1}}\right)^{\frac{\beta(\alpha+\sigma-1)}{(1-\sigma)(\beta-\alpha)}} \left[\sum_{\substack{j=1 \\ j\neq i}}^{n} \theta_j^{\frac{1-\sigma}{\beta+\sigma-1}} + \theta_i^{\frac{1-\sigma}{\beta+\sigma-1}} \left(1 - \frac{\alpha}{\beta}\theta_i\right)\right]. \tag{8}$$

To study the evolutionary stability of other-regarding preferences, we employ the indirect evolutionary approach pioneered by Güth and Yaari (1992). All players choose effort levels that maximize their utility, and evolution ensures the survival of phenotypes with preference parameters that induce equilibrium behavior providing the highest level of material payoff.

3 Evolutionary dynamics

3.1 Adaptive dynamics

The evolution of the preference trait θ under mutation and selection can be analyzed in a relatively straightforward way using adaptive dynamics (Dieckmann and Law, 1996;

---

[6] All the proofs of the results are given in the appendices.

[7] When $\sigma \in (1, +\infty)$, there also exists an (uninteresting) corner equilibrium, which involves zero effort by all players. The mathematical details of all these additional equilibria are given in the proof of Proposition 1 in Appendix A.

Geritz et al., 1998). The key notion of this approach is that of *invasion fitness*, which is given by the growth rate of a rare mutant characterized by preference θ' in a resident population that is monomorphic for θ. Two key assumptions of the adaptive dynamics framework are that mutants are phenotypically close to residents (i.e., $|\theta' - \theta| \ll 1$, "weak selection"), and that a rare mutant that is able to invade proceeds to get fixated in the population, thus changing the resident trait of the population to that of the mutant ("invasion implies substitution").

The invasion fitness of a mutant with preference θ' in a resident population monomorphic for θ is given by the mutant–resident payoff difference

$$f(\theta', \theta) = \Pi^*(\theta', \theta) - \Pi^*(\theta, \theta), \tag{9}$$

where $\Pi^*(\theta', \theta)$ is the equilibrium payoff of a mutant with preference trait θ' in a resident population monomorphic for θ, while $\Pi^*(\theta, \theta)$ is the equilibrium payoff of a resident player.

As we mentioned above, adaptive dynamics assumes that θ' is close to θ. In this case, invasion fitness $f(\theta', \theta)$ can be approximated by the first terms of its Taylor series expansion about $\theta' = \theta$, that is, by

$$f(\theta', \theta) = \left.\frac{\partial f}{\partial \theta'}\right|_{\theta'=\theta} (\theta' - \theta) + \frac{1}{2} \left.\frac{\partial^2 f}{\partial \theta'^2}\right|_{\theta'=\theta} (\theta' - \theta)^2 + \mathrm{h.o.t}, \tag{10}$$

where $\mathrm{h.o.t}$ denote higher order terms in the expansion.

An important quantity in this expression is the *selection gradient*, that is $D(\theta) \equiv \partial f / \partial \theta' \left.\right|_{\theta'=\theta}$. As long as $D(\theta) \neq 0$, the first term of Eq. (10) dominates other terms in the expansion and selection is directional, with the resident trait increasing or decreasing over evolutionary time. When $D(\theta) > 0$ holds, mutants with $\theta' > \theta$ have a positive invasion fitness and are thus able to invade and replace the resident population, making θ increase over evolutionary time. On the other hand, if $D(\theta) < 0$ holds, mutants with $\theta' < \theta$ invade and replace, making θ decrease over evolutionary time.

A *singular point* (or *singular trait*) θ^* is one at which the selection gradient vanishes, that is, one at which $D(\theta^*) = 0$ holds. At a singular point θ^*, the invasion fitness of a mutant with phenotype $\theta' \neq \theta^*$ depends on the second term of Eq. (10) evaluated at $\theta = \theta^*$. An important insight of adaptive dynamics is that it distinguishes between two stability criteria for singular points. The first one is *convergence stability*, which answers to the question of whether or not the singular point is attainable by evolution under a recurrent flow of mutations of small phenotypic effect. A singular point θ^* is *convergence stable* (or an *attractor*) if $dD/d\theta \left.\right|_{\theta=\theta^*} < 0$ holds. Conversely, if $dD/d\theta \left.\right|_{\theta=\theta^*} > 0$ holds, the singular point θ^* is said to be *convergence unstable* (or a *repeller*). The second criterion for stability is *evolutionary stability*. A singular point θ^* is *evolutionarily stable* if residents expressing θ^* cannot be invaded by phenotypically close mutants. An *evolutionarily stable* singular point θ^* must then be a local maximum of the invasion fitness $f(\theta', \theta^*)$, that is,

$\partial^2 f/\partial\theta'^2\,|_{\theta'=\theta=\theta^*} < 0$ needs to hold. If this inequality is reversed, θ^* can be invaded by mutants with parameter values on either side of the singular point θ^*, and is said to be *invasible*.

Crucially, convergence stability does not imply evolutionary stability, and vice versa. If a singular point is both convergence stable and evolutionarily stable, it is called a *continuously stable preference trait* (CSPT).[8] In this case, directional selection moves the population towards θ^*, and it remains monomorphic at such value. On the other hand, if a convergence stable singular point is invasible, directional selection first moves the population towards θ^*, at which point selection becomes disruptive leading to the splitting of the initially monomorphic population into distinct and diverging clusters, thus becoming polymorphic. In this case, θ^* is called an *evolutionary branching point*.

Much previous work has studied the adaptive dynamics of traits involved in social interactions modeled as games, but has so far focused on *strategy evolution*. In the following, we apply the framework of adaptive dynamics to study *preference evolution* for the case of the game of public good production presented in Section 2.

3.2 Preference evolution

Again, we focus on situations where the interior Nash equilibrium given by (7), in Proposition 1, holds. Consider a resident population characterized by the other-regarding preference trait θ while rare mutants are characterized by θ'. Substituting $\theta_i = \theta'$ for the mutant and $\theta_j = \theta$ for $j \neq i$ (i.e. each of the $n-1$ residents) in Eq. (8), we find that the payoff of a mutant with preference θ' in a resident population with preference θ is given by

$$\begin{aligned}\Pi^*(\theta',\theta) = {}& \alpha^{\frac{\alpha}{\beta-\alpha}} \left(\frac{1}{n}\right)^{\frac{\alpha\beta}{(\beta-\alpha)(1-\sigma)}} \left[\theta'^{\frac{1-\sigma}{\beta+\sigma-1}} + (n-1)\theta^{\frac{1-\sigma}{\beta+\sigma-1}}\right]^{\frac{\beta(\alpha+\sigma-1)}{(1-\sigma)(\beta-\alpha)}} \\ & \times \left[(n-1)\theta^{\frac{1-\sigma}{\beta+\sigma-1}} + \theta'^{\frac{1-\sigma}{\beta+\sigma-1}}\left(1 - \frac{\alpha}{\beta}\theta'\right)\right]. \end{aligned} \tag{11}$$

Its invasion fitness $f(\theta',\theta)$ is given by Eq. (9). The selection gradient can thus be written as

$$D(\theta) = \partial f/\partial\theta'|_{\theta'=\theta} = \frac{\alpha^{\frac{\beta}{\beta-\alpha}}\left[n(\beta+\sigma-1) - \theta(n(\beta-\alpha)+\alpha+\sigma-1)\right]\theta^{\frac{2\alpha-\beta}{\beta-\alpha}}}{n^{\frac{2\beta-\alpha}{\beta-\alpha}}(\beta-\alpha)(\beta+\sigma-1)}. \tag{12}$$

Solving $D(\theta^*) = 0$ for θ^*, we have the following result.

---

[8] We introduce the term *continuously stable preference trait* (CSPT) in reference to a *continuously stable strategy* (CSS). The notion of CSS originates from Eshel and Motro (1981) and Eshel (1983), later incorporated into adaptive dynamics as a convergence-stable ESS. The term CSPT indicates that dynamic evolution does not apply directly to strategies or behaviors, but indirectly to preference traits.

Proposition 2. *The adaptive dynamics is characterized by a unique singular point* $\theta^* \in (0, n)$*, hence* $\lambda^* \in (-1/(n-1), 1)$*, given by*

$$\theta^* = \frac{n(\beta + \sigma - 1)}{n(\beta - \alpha) + \alpha + \sigma - 1} \Rightarrow \lambda^* = \frac{\alpha + \sigma - 1}{n(\beta - \alpha) + \alpha + \sigma - 1}, \tag{13}$$

with

$$\lambda^* \gtreqless 0 \Leftrightarrow \sigma \gtreqless 1 - \alpha \Leftrightarrow \frac{\partial^2 V_i(\mathbf{e})}{\partial e_i \partial e_j} \gtreqless 0 \text{ for } i, j = 1, \ldots, n, \text{ and } i \neq j. \tag{14}$$

Finally, λ^* *is always increasing both in* α *and* σ*, and decreasing (resp. increasing) both in* β *and in group size* n *for* $\sigma \geq 1 - \alpha$ *(resp.* $\sigma \leq 1 - \alpha$*).*

It is worth noting at this stage that the private provision of a public good is typically subject to a *free-rider* or *collective action* problem. In a population of egoistic players, i.e. $\theta_i = 1$ for $i = 1, \ldots, n$, each player contributes to the public good up to the point where the marginal cost of their effort equals their marginal benefit from the public good, without taking into account that their contribution also benefits the other members of the group. The failure to internalize this positive externality implies that aggregate effort is suboptimal.[9]. Thus, a positive value of λ^* mitigates the collective action problem that causes suboptimal provision of the public good, while a negative value exacerbates it. Proposition 2 states that if the degree of complementarity between individual efforts is sufficiently high, i.e., if $\sigma > 1 - \alpha$ holds, then the singular point is characterized by benevolence ($\lambda^* > 0$). Conversely, if $\sigma < 1 - \alpha$ holds, the singular point is characterized by malevolence ($\lambda^* < 0$). Preferences corresponding to the paradigm of *Homo economicus* (i.e., egoistic players with $\lambda^* = 0$) is a singular point only in the knife-edge case where $\sigma = 1 - \alpha$.

Proposition 2 also states that the sign of λ^* (and hence, whether or not the singular trait expresses benevolence or malevolence) does not depend on the size of the group or the elasticity of the cost function of individual effort, but solely on the sign of the cross derivative of V_i with respect to e_i and e_j. In the economic literature, following Bulow et al. (1985), the sign of this cross-derivative determines the nature of *strategic interactions* between players. When it is negative (resp. positive), players' actions are called strategic substitutes (resp. complements) because the optimal response of a player to higher effort by another player is to reduce (resp. increase) effort.[10] Hence, when players' efforts are

---

[9] Moreover, the collective action problem is exacerbated as group size increases, since the gap between the optimal aggregate effort and that resulting from Nash equilibrium becomes greater. Indeed, using Eq. (7) and setting $\theta_i = 1$ for $i = 1, \ldots, n$, we have $e^* = (\alpha/n)^{1/(\beta-\alpha)}$, while the optimal level of individual effort is independent of group size and given by $e^\dagger = \alpha^{1/(\beta-\alpha)}$.

[10] The optimal response of player i to another player's effort is implicitly given by the first-order condition given by Eq. (A.6) in Appendix A. This Appendix also shows that $\partial^2 V_i(\mathbf{e})/e_i^2 < 0$. By the implicit function theorem, the optimal response of player i to player j' effort is thus given by $\partial e_i/\partial e_j = -[(\partial^2 V_i(\mathbf{e})/\partial e_i \partial e_j)/(\partial^2 V_i(\mathbf{e})/\partial e_i^2)]$. The sign of $\partial e_i/\partial e_j$, for $i \neq j$, is given by the sign of its numerator, that is by the sign of the cross derivative of V_i with respect to e_i and e_j.

strategic substitutes, the preferences at the singular point are characterized by malevolence, leading to lower contributions than if the population consisted of egoists. Indeed, a malevolent individual can invade a population of egoistic players because, by exerting less effort than a resident player, he induces his partners to exert more. The resulting savings in effort costs outweigh the loss in the public good due to his reduced contribution. Conversely, when players' efforts are strategic complements, the preferences at the singular point are characterized by benevolence, leading to higher contributions than if the population consisted of egoists. Indeed, a benevolent individual, thus exhibiting cooperative behavior, can invade a population of malevolent players since, by exerting more effort, he encourages his partners to do the same. The resulting increase in public good production offsets the additional effort cost. In sum, under complete information, biased preferences toward benevolence or malevolence, by inducing behavior that differs from that implied by purely selfish preferences, can replace a population of *Homo economicus*. This occurs because such preferences alter the behavior of selfish individuals when they interact with biased types rather than with other *Homo economicus*. In the special case where $\sigma = 1 - \alpha$, biased preferences still generate behavior distinct from that of *Homo economicus*, but they do not affect the behavior of partners. As a result, biased preferences cannot invade a population of selfish players.[11]

Finally, we observe that λ^* is monotonically increasing with the public-good satisfaction scale parameter α and with σ, the parameter capturing the degree of complementarity among individual efforts. Moreover, λ^* is increasing in both the elasticity of the cost function β and group size n whenever $\lambda^* < 0$, that is, when $\sigma \leq 1 - \alpha$. By contrast, when $\sigma > 1 - \alpha$, λ^* is positive and decreasing in both β and n. In other words, as either the elasticity of effort costs or group size increases, the singular point λ^* converges toward zero, corresponding to the *Homo economicus* paradigm of purely selfish behavior.

We turn now to the stability of the singular point θ^* (hence λ^*) according to the two criteria recognized in adaptive dynamics. To begin with, we consider convergence stability. The derivative of the selection gradient at the singular point satisfies

$$dD/d\theta\big|_{\theta=\theta^*} = -\frac{\alpha^{\frac{\beta}{\beta-\alpha}}\left[n(\beta-\alpha)+\alpha+\sigma-1\right]^{\frac{2\beta-3\alpha}{\beta-\alpha}}}{n^3(\beta-\alpha)(\beta+\sigma-1)^{\frac{2\beta-3\alpha}{\beta-\alpha}}}, \tag{15}$$

which is negative. Thus, θ^* is always convergence stable. Next, we verify under which condition θ^* is a continuously stable trait or an evolutionary branching point. To do so, we evaluate the second partial derivative of invasion fitness at the singular point

$$\partial^2 f/\partial\theta'^2\big|_{\theta'=\theta=\theta^*} = \frac{\alpha^{\frac{\beta}{\beta-\alpha}}\left[n(\beta-\alpha)+\alpha+\sigma-1\right]^{\frac{\beta-2\alpha}{\beta-\alpha}}\ K(\alpha,\beta,\sigma,n)}{n^4(\beta-\alpha)(\beta+\sigma-1)^{\frac{3\beta-4\alpha}{\beta-\alpha}}}, \tag{16}$$

---

[11] For a general discussion on the link between the nature of strategic interactions and the parameters of evolutionarily stable social preferences with assortative matching, see (Alger and Weibull, 2012). For a study of evolutionarily stable preferences in a conflict model with both intragroup and inter-group strategic interactions, see (Cheikbossian, 2021b,a).

where

$$\begin{aligned}K(\alpha,\beta,\sigma,n) \equiv & -\sigma^2 - \sigma\left[n^2(\beta-\alpha)+n\beta+\alpha-2\right] \\ & + n\beta[n+1-\alpha-n(\beta-\alpha)] - \alpha(n^2-1) - 1. \end{aligned} \tag{17}$$

We have the following result.

Lemma 1. *The sign of* $\partial^2 f/\partial\theta'^2|_{\theta'=\theta=\theta^*}$ *depends on the sign of* $K(\alpha,\beta,\sigma,n)$ *given by* (17)*, which is concave in* σ*.* $K(\alpha,\beta,\sigma,n)=0$ *has thus two solutions in* σ*. The first one is always negative. The second one, denoted by* $\bar{\sigma}(\alpha,\beta,n)$*, is given by*

$$\bar{\sigma}(\alpha,\beta,n) = \frac{1}{2}\left\{-\left[n^2(\beta-\alpha)+n\beta+\alpha-2\right]+\Gamma(\alpha,\beta,n)\right\}, \tag{18}$$

where

$$\Gamma(\alpha,\beta,n) = \sqrt{(n-1)[n(\beta-\alpha)+\alpha][\beta n(n+3)-\alpha(n+1)^2]}. \tag{19}$$

Furthermore, $\bar{\sigma}(\alpha,\beta,n)$ *is decreasing in* α*,* β *and* n*. It also satisfies* $\bar{\sigma}(\alpha,\beta,n) < 1-\alpha$ *for any* $\alpha \in (0,1)$*,* $\beta \geq 1$ *and* $n \geq 2$*.*

This result shows that even if players have concave utility functions in their own actions, this does not imply that their resulting material payoffs, determined by equilibrium actions that depend on social preference traits, will themselves be concave in those traits. Put differently, there is no general guarantee that social preferences maximizing evolutionary fitness will exist in economic games, even in relatively simple environments. This can give rise to evolutionary branching in social preferences. We have the following result.

Proposition 3. *Let* $\bar{\sigma}(\alpha,\beta,n)$ *be defined by* (18) *and* (19)*. Define now* $\beta^*(\alpha,n) > 1$ *such that* $\bar{\sigma}(\alpha,\beta^*(\alpha,n),n)=0$*. Then, the adaptive dynamics can be classified as follows:*

- *(i) If either* $\beta > \beta^*(\alpha,n)$*, or* $\beta \in [1,\beta^*(\alpha,n)]$ *and* $\sigma > \bar{\sigma}(\alpha,\beta,n) \geq 0$*, then the singular point* λ^**, given by* (13)*, is a continuously stable preference trait (CSPT). In this case, the population exhibits malevolence (*$\lambda^* \leq 0$*) when* $\sigma \leq 1-\alpha$*, and benevolence (*$\lambda^* \geq 0$*) when* $\sigma \geq 1-\alpha$*.*
- *(ii) If* $\beta \in [1,\beta^*(\alpha,n))$ *and* $0 \leq \sigma < \bar{\sigma}(\alpha,\beta,n)$*, then the singular point* $\lambda^* < 0$ *is an evolutionary branching point.*

A monomorphic population where all individuals have the same social preferences and therefore the same behaviors is an evolutionarily stable equilibrium if the elasticity of the cost function of effort is sufficiently large—i.e. if $\beta > \beta^*(\alpha,n) > 1$—or if the degree of complementarity between individual efforts in joint production is sufficiently high—i.e. $\sigma > \bar{\sigma}(\alpha,\beta,n)$. In this case, the population is benevolent (resp. malevolent) if the degree of effort complementarity is larger (resp. smaller) than the degree of decreasing returns to scale in satisfaction obtained by consuming the public good, which is given by $1-\alpha$. If, on the contrary, the elasticity of the cost function of effort is relatively small—i.e. if $\beta \in [1, \beta^*(\alpha, n))$—and if individual efforts are too substitutable—i.e. $\sigma < \bar{\sigma}(\alpha, \beta, n)$—, then the population converges towards a situation where individuals value their partners' material payoffs negatively, before separating into distinct populations, thus characterizing a polymorphic population with heterogeneous preference traits.

To get further insight, let us now consider the limiting case of constant marginal cost (i.e. $\beta = 1$). This is the reference assumption in the economics literature on public good games when there are decreasing returns to scale in the satisfaction obtained by public good consumption (i.e. $\alpha < 1$). In this case, we have $\bar{\sigma}(\alpha, 1, n) > 0$. Then, combining the results of Propositions 2 and 3, we can conclude that if the degree of complementarity between individual efforts in joint production is sufficiently high (i.e. $\sigma > \bar{\sigma}(\alpha, 1, n)$), a larger group size attenuates the continuously stable degree of benevolence or malevolence (depending on whether the degree of complementarity is sufficiently large relative to $1-\alpha$).[12] If the degree of complementarity between individual efforts in joint production is relatively low (i.e. $\sigma < \bar{\sigma}(\alpha, 1, n)$), evolutionary branching of social preferences occurs. Indeed, when efforts are highly substitutable, a monomorphic population at the singular point becomes unstable to invasion by agents who place either greater or lesser weight on others' gains than the resident population. This is because any reduction in one individual's contribution can be readily compensated by increased contributions from partners, allowing the same level of public goods to be produced.

Interestingly, from Lemma 1, we also have that a larger group size makes evolutionary branching less likely because it decreases the threshold value of the degree of complementarity between individual efforts above which preference evolution converges towards a continuously stable preference trait. Again, evolutionary branching can occur only when $\sigma < 1 - \alpha$, in which case the singular point is negative ($\lambda^* < 0$), corresponding to malevolent preferences. Recall moreover that the degree of malevolence decreases with group size. That is, as group size increases, preferences evolve toward selfishness. Intuitively, the larger the group, the weaker the impact of biased preferences on others' behavior, making a population of players with preferences close to selfishness less likely to be destabilized by mutants with alternative biased preferences.

3.3 Individual-based simulations

To complement our analytical results, we simulate the evolution of the preference trait λ in a finite population using an individual-based Wright–Fisher process. In each generation, N individuals are matched uniformly at random into groups of size n, material payoffs are computed from the group composition using Eq. (8), and these payoffs are transformed into non-negative fecundities that determine reproductive success in the

---

[12] One can also observe that λ^* given by Eq. (13) approaches 0 as group size n tends to infinity.

Wright–Fisher process. Mutations occur independently across offspring with probability μ, with Gaussian effect size of standard deviation s, and traits are constrained to remain in the admissible interval ensuring $\theta > 0$. We initialize the population near $\lambda = 0$, run the process for T generations, and save periodic snapshots of the full population to visualize convergence and branching.

Figure 1 illustrates the role of the complementarity parameter σ for $\alpha = 1/2$, $\beta = 1$, and $n = 2$. These parameter values imply $1 - \alpha = 1/2$, so the singular trait is benevolent when $\sigma > 1/2$ and malevolent when $\sigma < 1/2$. In panel A, where $\sigma = 0.7$, the population converges toward a positive singular trait $\lambda^* > 0$, consistent with a continuously stable preference trait. In panel B, where $\sigma = 0.3$, the singular point is negative but remains evolutionarily stable, and the population converges to a monomorphic malevolent state. In panel C, where $\sigma = 0.1$, the singular point is still convergence stable but becomes invasible, so the initially monomorphic population first approaches λ^* and then splits into distinct coexisting clusters. The three panels, therefore, reproduce the three qualitative regimes highlighted by Proposition 3: convergence to benevolence, convergence to malevolence, and evolutionary branching under sufficiently high substitutability of efforts.

Figure 2 illustrates the role of group size. We fix $\alpha = 1/2$, $\beta = 1$, and $\sigma = 0.16$, and compare the cases $n = 2$ and $n = 4$. For $n = 2$, the population evolves toward the singular point and then branches into multiple clusters. By contrast, when $n = 4$, branching disappears and the population converges toward a monomorphic distribution centered around a negative singular trait. This comparison is in line with Lemma 1 and Proposition 3, which show that the threshold $\bar{\sigma}(\alpha, \beta, n)$ decreases with n. Hence, for fixed α, β, and σ, increasing group size reduces the parameter range under which disruptive selection arises.

Taken together, the simulations confirm the main theoretical predictions of the model. High complementarity leads to convergence toward a homogeneous population, with benevolent or malevolent preferences depending on whether σ lies above or below $1 - \alpha$. By contrast, when efforts are sufficiently substitutable, the singular point becomes a branching point and heterogeneous social preferences emerge. The comparison across group sizes further shows that larger groups dampen this tendency toward polymorphism, pushing the evolutionary outcome toward monomorphism and preferences closer to egoism.

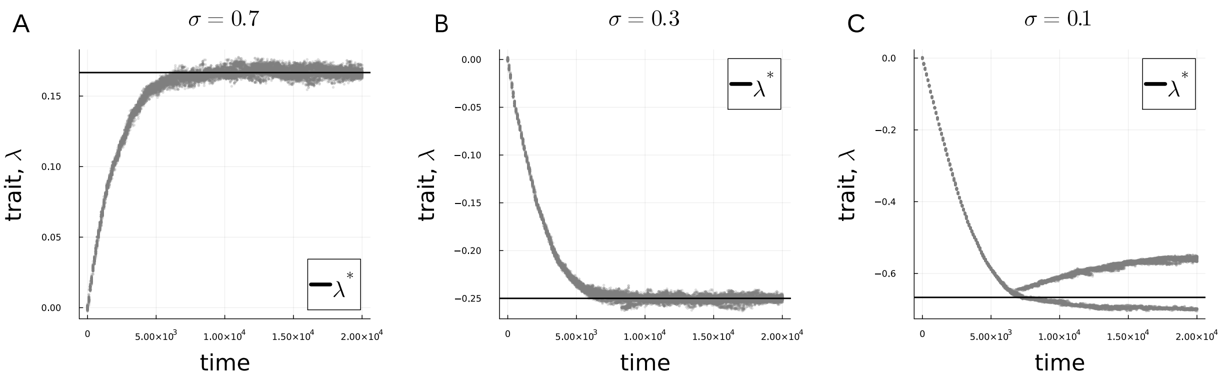

Figure 1: Evolution of the preference trait λ for different values of the complementarity parameter σ. The figure reports individual-based Wright–Fisher simulations for $\alpha = 0.5$, $\beta = 1$, and $n = 2$ (implying $\bar{\sigma} \approx 0.186$), with population size $N = 5000$, mutation probability $\mu = 0.01$, mutation standard deviation $s = 0.001$, and $T = 2 \times 10^4$ generations. Gray dots show a random sample of 100 individuals from the population, recorded every 100 generations. The solid horizontal line indicates the singular trait λ^* predicted by Eq. (13). **A:** $\sigma = 0.7$, for which the population converges to a monomorphic benevolent state $\lambda^* \approx 0.17$. **B:** $\sigma = 0.3$, for which the population converges to a monomorphic malevolent state $\lambda^* = -0.25$. **C:** $\sigma = 0.1$, for which the population first approaches the singular point $\lambda^* \approx -0.67$ and then branches into distinct coexisting clusters.

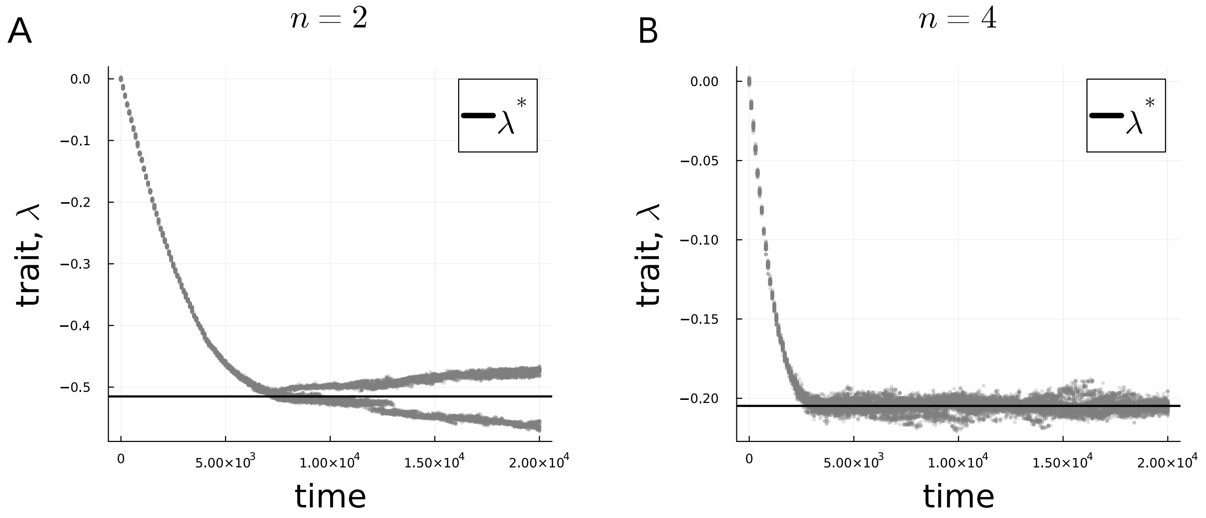

Figure 2: Evolution of the preference trait λ for different group sizes. The figure reports individual-based Wright–Fisher simulations for $\alpha = 0.5$, $\beta = 1$, and $\sigma = 0.16$. Other simulation parameters are as in Fig. 1. The solid horizontal line indicates the analytical value of the singular trait λ^*. **A:** $n = 2$, for which $\bar{\sigma} \approx 0.186$. The population branches after reaching the singular point $\lambda^* \approx -0.52$. **B:** $n = 4$, for which $\bar{\sigma} \approx 0.141$. Branching disappears, and the population converges to a monomorphic malevolent state $\lambda^* \approx -0.20$.

4 Discussion

4.1 Preference versus strategy evolution

The literature on adaptive dynamics that has developed in biology logically focuses on the genetically inherited strategies of players. In their influential paper, Doebeli et al. (2004) used adaptive dynamics to show how populations can evolve from full cooperation to coexistence of cooperators and defectors, or even to full defection, depending on the payoff structure of the public goods game. In our specific model of public goods production, the material payoff for each actor is strictly concave with respect to their level of effort. There is a unique strict Nash equilibrium that is both convergent stable and evolutionarily stable under adaptive dynamics.[13] Evolutionary branching cannot occur at the strategy level in our model.

Doebeli et al. (2004) show that evolutionary branching with coexistence of cooperators and cheaters is only under certain conditions related to the marginal utility of the public good and the marginal cost of individual contribution. Using their notation, where x denotes the resident strategy and y denotes the mutant strategy, the payoff of a y-strategist interacting with a x-strategist is: $\pi(y,x) = G(x+y) - c(y)$. Note that the benefit of the public good, i.e. $G(x+y)$ depends on the sum of individual efforts, which are thus assumed to be perfectly substitutable. The invasion fitness of a mutant y-strategist is given by $f(y,x) = \pi(y,x) - \pi(x,x)$, and the selection gradient by $D(x) = \partial f / \partial y\,|_{y=x} = G'(2x) - c'(x)$. The singular strategy x^* of the adaptive dynamics is such that $D(x^*) = G'(2x^*) - c'(x^*) = 0$. Under conditions of existence, x^* is convergence stable if $dD/dx\,|_{x=x^*} = 2G''(2x^*) - c''(x^*) < 0$, and x^* is evolutionarily stable if $\partial^2 f / \partial y^2\,|_{y=x=x^*} = G''(2x^*) - c''(x^*) < 0$. Under the standard assumptions that the benefit function is concave and the cost function is convex, both conditions hold and the singular strategy x^* is both convergence stable and evolutionarily stable. The singular strategy x^* is an evolutionarily branching point only if $2G''(2x^*) < c''(x^*) < G''(2x^*) < 0$, that is, only if both cost and production functions are concave, near to the singular point.[14] Although the assumption of diminishing marginal benefits is clearly relevant, the same cannot be said of the assumption of diminishing marginal costs. Doebeli et al. (2004) give the example of enzyme production machinery to illustrate the case of diminishing marginal costs. However, as far as human activities are concerned, it is very difficult to find a concrete example with such technical characteristics. Even if learning effects can reduce the marginal cost of effort at the start of a collective project, it is clear that it must then rapidly increase, if only because humans have a limited endowment of time. In the present analysis, we consid-

---

[13] As shown in Appendix A, the second derivative of the utility function of player i with respect to e_i is strictly negative, implying that her material payoff is strictly concave in her effort level.

[14] This can be easily generalized to a group size $n \geq 2$. In this case, x^* is an evolutionary branching point if $nG''(nx^*) < c''(x^*) < G''(nx^*)$ holds. This means that, in contrast to our result on the evolution of preferences, the range of parameters leading to evolutionary branching increases with group size.

ered the more realistic case of increasing marginal cost of efforts and decreasing marginal benefits. In this case, evolutionary branching cannot occur at the strategy level. Yet, it can occur at the level of individual preferences, which determine their equilibrium behavior.

4.2 Linear-quadratic payoffs

After completing this research, we became aware of a recent article by Lehmann and Mullon (2025) developing a general analysis of adaptive polymorphisms in continuous traits. Among other applications, they study the evolution of other-regarding preferences in a two-player game with linear-quadratic material payoffs, following Bester and Güth (1998) and Heiftez et al. (2007a). In this setting, the material payoff of a player with action x interacting with a player with action y is $\Pi(x, y) = x(ky + m - x)$, where k captures the effect of the partner's effort on one's own payoff. The sign of k determines the nature of strategic interactions: efforts are strategic substitutes when $k < 0$ and strategic complements when $k > 0$. Each player's utility adds a weighted term $\tau\Pi(y, x)$ or $\gamma\Pi(x, y)$, where the weight is the evolving preference trait.

Applying the adaptive dynamics machinery to this model (see Appendix E for details), we find a unique singular preference trait $\gamma^* = k/(2 - k)$, which is malevolent when $k < 0$ and benevolent when $k > 0$. This follows the same qualitative pattern as our Proposition 2. However, the stability properties differ markedly from our model. Under the economically relevant restriction $k \in (-1, 1)$, the singular point is both convergence stable and evolutionarily stable (except at the degenerate benchmark $k = 0$), and hence a continuously stable preference trait. Evolutionary branching requires $k < -2$: the negative effect of the partner's effort on one's own payoff must be more than twice as strong as the own quadratic cost of effort. Such a parameter configuration lies outside the range considered by Bester and Güth (1998), Heiftez et al. (2007a), and related contributions, which is why these analyses identify an ESS for social preferences.

The contrast with our results is instructive. In the linear-quadratic model, the payoff structure is too simple—individual contributions enter linearly in the partner's payoff—to generate the non-concavity of fitness in the preference parameter that is needed for evolutionary branching under realistic conditions. Our CES public goods game, by contrast, allows the degree of substitutability among contributions to vary continuously. It is this richer production structure that enables evolutionary branching within empirically plausible parameter ranges, even when individual utility remains concave in effort.

5 Conclusions

A substantial body of work has examined the evolution of social behavior. Approaches grounded in cultural evolution have highlighted how social traits, including cooperative dispositions, can be transmitted and maintained through learning and social influence

(Cavalli-Sforza and Feldman, 1981; Boyd and Richerson, 1985, 2005), and how selection may act on preferences that support reciprocity and norm enforcement (Bowles and Gintis, 2011). Empirical studies further document substantial cross-cultural variation in cooperative behavior and social preferences (Henrich et al., 2005). The present results complement these approaches by providing a formal mechanism through which diversity in preferences may arise endogenously.

While adaptive dynamics has traditionally been applied to strategies or traits directly expressed in behavior, our results indicate that the preference structure underlying decision-making may itself diversify under evolutionary dynamics. Behavioral diversity, in this framework, does not stem from exogenous heterogeneity or noise, but from selection acting on preferences that determine optimal behavior in a given strategic environment. In particular, we show that evolutionary branching can occur in a standard public goods game when individual contributions are sufficiently substitutable in the production of the public good. This suggests that diversification in preference traits may provide a unifying explanation for the coexistence of qualitatively different behavioral types in strategic environments such as team production. Such diversity is documented, for example, in small-scale societies such as the Hadza, where individuals differ in foraging strategies and food-sharing behavior (Smith et al., 2022; Stibbard-Hawkes et al., 2022).

Although the conditions for evolutionary branching can be characterized, the evolutionary dynamics following branching remain analytically intractable in our framework. Our simulations suggest that diversification can lead to persistent polymorphism in preferences, but do not yet allow a full characterization of its stability or long-term dynamics. This limitation reflects a broader challenge in adaptive dynamics, where the identification of branching points does not necessarily imply convergence to a stable dimorphism (or polymorphism) [15]. In the present context, the mapping from preferences to behavior introduces an additional layer of complexity, as selection acts indirectly through equilibrium outcomes of strategic interactions. Future work should aim to better characterize the evolutionary dynamics beyond branching, for example by incorporating additional structure such as population subdivision or assortative interactions, in order to reduce analytical complexity by constraining the range of strategic interactions. With the support of individual-based simulations, it could provide further insight into the conditions under which diversification in preferences leads to stable coexistence.

CRediT authorship contribution statement

Guillaume Cheikbossian: Conceptualization, Methodology, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Funding acquisition; **Jorge Peña**: Conceptu-

---

[15] For instance, Doebeli et al. (2004), in supporting material, show that even in relatively simple settings with quadratic cost and benefit functions, the emergence and stability of dimorphism depend on restrictive parameter conditions.

alization, Methodology, Software, Writing - Original Draft, Writing - Review & Editing, Visualization, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of generative AI use

During the preparation of this work the authors used DeepL and ChatGPT to assist with English-language editing in parts of the text. The authors also used Claude to assist with the verification of mathematical arguments and calculations, and with the writing of the code for the computer simulations. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Acknowledgments

We thank seminar participants at TREE (Université de Pau), GAEL (Université Grenoble Alpes), CEE-M (Université de Montpellier) and participants at the 2025 ASSET Meeting in Rabat. We acknowledge financial support from the French National Research Agency (ANR-22-CE26-0014-01 and ANR-17-EURE-0010). The code for the simulations is available at `https://github.com/jorgeapenas/PreferenceBranching`.

A Proof of Proposition 1

Existence. We first show the existence of a pure-strategy Nash equilibrium. First, each player can invest at most her endowment y in the collective good, so the strategy space of each player is the compact interval $S = [0, y]$. We now show that each player's maximization problem is strictly concave. The first derivative of $V_i(\mathbf{e})$ with respect to e_i is given by

$$\frac{\partial V_i(\mathbf{e})}{\partial e_i} = \alpha\theta_i \left[G(\mathbf{e})\right]^{\alpha-1} \frac{\partial G(\mathbf{e})}{\partial e_i} - e_i^{\beta-1}. \tag{A.1}$$

Using Eq. (1), we have

$$\frac{\partial G(\mathbf{e})}{\partial e_i} = \frac{1}{n}\left(\frac{1}{n}\sum_{j=1}^{n} e_j^{1-\sigma}\right)^{\frac{1}{1-\sigma}-1} e_i^{-\sigma} = \frac{1}{n}\left[G(\mathbf{e})\right]^{\sigma} e_i^{-\sigma}. \tag{A.2}$$

The first-order condition is thus given by

$$\frac{\partial V_i(\mathbf{e})}{\partial e_i} = \frac{\alpha}{n}\theta_i \left[G(\mathbf{e})\right]^{\alpha+\sigma-1} e_i^{-\sigma} - e_i^{\beta-1} \leq 0, \quad \text{with equality for} \quad e_i > 0. \tag{A.3}$$

The second derivative of $V_i(\mathbf{e})$ with respect to e_i is given by

$$\frac{\partial^2 V_i(\mathbf{e})}{\partial e_i^2} = \frac{\alpha}{n}\theta_i \left[G(\mathbf{e})\right]^{\alpha+2\sigma-2} e_i^{-\sigma-1}\left[\frac{1}{n}(\alpha+\sigma-1)e_i^{1-\sigma} - \sigma\left[G(\mathbf{e})\right]^{1-\sigma}\right] - (\beta-1)e_i^{\beta-2}. \tag{A.4}$$

Using Eq. (1), we have

$$\frac{\partial^2 V_i(\mathbf{e})}{\partial e_i^2} = -\frac{\alpha}{n^2}\theta_i \left[G(\mathbf{e})\right]^{\alpha+2\sigma-2} e_i^{-\sigma-1}\left[(1-\alpha)e_i^{1-\sigma} + \sigma\sum_{\substack{j=1\\ j\neq i}}^{n} e_j^{1-\sigma}\right] - (\beta-1)e_i^{\beta-2}, \tag{A.5}$$

which is strictly negative for strictly positive effort levels, $\alpha \in (0,1)$, $\beta \geq 1$ and $\sigma \in [0,1) \cup (1,\infty)$. Hence, $V_i(\mathbf{e})$ is strictly concave and continuous in e_i, which guarantees the existence of a pure-strategy Nash equilibrium in efforts. In this equilibrium, the first-order condition given by (A.3) is both necessary and sufficient for characterizing player i's best-response function.

Uniqueness. Suppose $\beta + \sigma - 1 \neq 0$. Then, using Eq. (A.3), the first-order interior condition $\partial V_i(\mathbf{e})/\partial e_i = 0$, can be rewritten as

$$e_i = \left(\frac{\alpha\theta_i}{n}\right)^{\frac{1}{\beta+\sigma-1}} \left[G(\mathbf{e})\right]^{\frac{\alpha+\sigma-1}{\beta+\sigma-1}}. \tag{A.6}$$

Raise to the power $1-\sigma$, sum over i and divide by n, we have

$$\frac{1}{n}\sum_{i=1}^{n} e_i^{1-\sigma} = G(\mathbf{e})^{1-\sigma} = \alpha^{\frac{1-\sigma}{\beta+\sigma-1}}\left(\frac{1}{n}\right)^{\frac{\beta}{\beta+\sigma-1}} \left[G(\mathbf{e})\right]^{\frac{(1-\sigma)(\alpha+\sigma-1)}{\beta+\sigma-1}} \sum_{i=1}^{n} \theta_i^{\frac{1-\sigma}{\beta+\sigma-1}}. \tag{A.7}$$

Solving in $G(\mathbf{e})$, we obtain

$$[G(\mathbf{e})]^{\omega} = \alpha^{\frac{1-\sigma}{\beta+\sigma-1}} \left(\frac{1}{n}\right)^{\frac{\beta}{\beta+\sigma-1}} \sum_{i=1}^{n} \theta_i^{\frac{1-\sigma}{\beta+\sigma-1}} \quad \text{with} \quad \omega \equiv \frac{(1-\sigma)(\beta-\alpha)}{\beta+\sigma-1}. \tag{A.8}$$

The mapping $G(\mathbf{e}) \to [G(\mathbf{e})]^{\omega}$ is continuous and strictly monotone on $(0, +\infty)$. It is strictly increasing for $\sigma \in [0,1)$ and strictly decreasing for $\sigma \in (1, +\infty)$. Hence, Eq. (A.8) has exactly one positive solution for $G(\mathbf{e})$. Substituting this unique $G(\mathbf{e})$ into Eq. (A.6) yields a unique strictly positive e_i for $i = 1, \ldots, n$, given by Eq. (7) in Proposition 1. Note also that if $\sigma \in (1, +\infty)$ and player i exerts 0 effort then, as mentioned in the text, we take the limit of Eq. (1), i.e. $G(\mathbf{e}) = 0$. In this case, $e_j^* = 0$ for $j = 1, \ldots, n$ are mutually best-responses, a corner equilibrium that we disregard.

Suppose now that $\beta + \sigma - 1 = 0$ holds, which implies $\beta = 1$ (constant marginal cost of effort) and $\sigma = 0$ (perfect substitutability). Using Eq. (1) and Eq. (A.3), the first-order condition becomes $\alpha(1/n)^{\alpha}\theta_i \left[\sum_{i=1}^{n} e_i\right]^{\alpha-1} \leq 1$, with equality for $e_i > 0$. Observe that this condition cannot hold with equality for any two players i and j, with different preference parameters (i.e. $\theta_i \neq \theta_j$). Let $\bar{\theta} = \max\{\theta_1, \ldots, \theta_n\}$ and $\mathcal{A} = \{i : \theta_i = \bar{\theta}\}$. If $|\mathcal{A}| = 1$, there exists a unique Nash equilibrium with $e_i^* = nG^* = \left[\alpha\bar{\theta}/n^{\alpha}\right]^{1/(1-\alpha)}$ and $e_j^* = 0$ for $j \neq i$. If $|\mathcal{A}| \geq 2$, there exists a continuum of Nash equilibria, all with the same level of public good production given by $\sum_{i \in \mathcal{A}} e_i^* = nG^* = \left[\alpha\bar{\theta}/n^{\alpha}\right]^{1/(1-\alpha)}$ and where $e_j^* = 0$ for $j \notin \mathcal{A}$. In these cases, the player(s) with the highest other-regarding preference trait assume full responsibility for the provision of the public good, whereas all other players free ride completely.

Finally, consider the other extreme case of perfect complementarity ($\sigma \to +\infty$), corresponding to the weakest-link technology (i.e., $G(\mathbf{e}) = \min\{e_1, \ldots, e_n\}$). Let $\underline{\theta} = \min\{\theta_1, \ldots, \theta_n\}$ and $\underline{G} = (\alpha\underline{\theta})^{1/(\beta-\alpha)}$. Then, there exists a continuum of Nash equilibria where all players exert the same effort level, i.e. $e_j^* = e^*$ for $j = 1, \ldots, n$, satisfying $0 \leq e^* \leq \underline{G}$. Intuitively, it is not beneficial for any player to exert more efforts than his partners because it increases his cost of effort without increasing the public good level. And the common equilibrium level of effort can be any value between 0 and $\underline{G}$ that maximizes the utility of the player with the smallest other-regarding preference trait.[16]

B Proof of Proposition 2

First, observe that the denominator of λ^* given by Eq. (13) is always positive. The sign of λ^* is thus given by the sign of its numerator, that is, by the sign of $\alpha + \sigma - 1$. Now using Eq. (A.3), the cross derivative of $V_i(\mathbf{e})$ with respect to e_i and e_j is given by

$$\frac{\partial^2 V_i(\mathbf{e})}{\partial e_i \partial e_j} = \frac{\alpha}{n}(\alpha + \sigma - 1)\theta_i \left[G(\mathbf{e})\right]^{\alpha+\sigma-2} \frac{\partial G(\mathbf{e})}{\partial e_j} e_i^{-\sigma}. \tag{B.1}$$

---

[16] For a rigorous proof that relies on "replacement functions", see Cornes and Hartley (2007).

Since $\partial G(\mathbf{e})/\partial e_j > 0$, the sign of $\partial^2 V_i(\mathbf{e})/\partial e_i \partial e_j$ is the same as the sign of $\alpha + \sigma - 1$.

Finally, the derivatives of λ^* with respect to α, σ, β and n are given by

$$\frac{\partial \lambda^*}{\partial \alpha} = \frac{n(\beta + \sigma - 1)}{[n(\beta - \alpha) + \alpha + \sigma - 1]^2} > 0, \tag{B.2}$$

$$\frac{\partial \lambda^*}{\partial \sigma} = \frac{n(\beta - \alpha)}{[n(\beta - \alpha) + \alpha + \sigma - 1]^2} > 0, \tag{B.3}$$

$$\frac{\partial \lambda^*}{\partial \beta} = -\frac{n(\alpha + \sigma - 1)}{[n(\beta - \alpha) + \alpha + \sigma - 1]^2} \gtreqless 0 \Leftrightarrow \sigma \lesseqgtr 1 - \alpha, \tag{B.4}$$

$$\frac{\partial \lambda^*}{\partial n} = -\frac{(\beta - \alpha)(\alpha + \sigma - 1)}{[n(\beta - \alpha) + \alpha + \sigma - 1]^2} \gtreqless 0 \Leftrightarrow \sigma \lesseqgtr 1 - \alpha. \tag{B.5}$$

C Proof of Lemma 1

We first demonstrate that $\bar{\sigma}(\alpha, \beta, n)$ is decreasing in α. We have

$$\frac{\partial \bar{\sigma}(\alpha, \beta, n)}{\partial \alpha} = \frac{(n-1)\left\{(n+1)\Gamma(\alpha, \beta, n) - [\beta n(n^2 + 2n - 1) - \alpha(n+1)(n^2 - 1)]\right\}}{2\Gamma(\alpha, \beta, n)}, \tag{C.1}$$

where $\Gamma(\alpha, \beta, n)$ is given by (19).

The sign of $\partial \bar{\sigma}(\alpha, \beta, n)/\partial \alpha$ is the same as the sign of its numerator. Let $P(\alpha, \beta, n) = (n+1)^2 \Gamma(\alpha, \beta, n)^2 - [\beta n(n^2 + 2n - 1) - \alpha(n+1)(n^2 - 1)]^2$. It simplifies to $P(\alpha, \beta, n) = -4n^2\beta^2 < 0$. It follows that $\partial \bar{\sigma}(\alpha, \beta, n)/\partial \alpha$ is strictly negative.

We first demonstrate that $\bar{\sigma}(\alpha, \beta, n)$ is decreasing in β. We have

$$\frac{\partial \bar{\sigma}(\alpha, \beta, n)}{\partial \beta} = \frac{n\left\{-(n+1)\Gamma(\alpha, \beta, n) + (n-1)[\beta n(n+3) - \alpha n(n+2) + \alpha]\right\}}{2\Gamma(\alpha, \beta, n)}, \tag{C.2}$$

where $\Gamma(\alpha, \beta, n)$ is given by (19).

The sign of $\partial \bar{\sigma}(\alpha, \beta, n)/\partial \beta$ is the same as the sign of its numerator. Let $M(\alpha, \beta, n) = (n-1)^2[\beta n(n+3) - \alpha n(n+2) + \alpha]^2 - (n+1)^2 \Gamma(\alpha, \beta, n)^2$. It simplifies to $M(\alpha, \beta, n) = -4(\beta - \alpha)n(n-1)[\beta n(n+3) - \alpha(n^2 + n - 2)]$, which is strictly negative for $\alpha \in (0, 1)$, $\beta \geq 1$ and $n \geq 2$. It follows that $\partial \bar{\sigma}(\alpha, \beta, n)/\partial \beta$ is strictly negative.

We now demonstrate that $\bar{\sigma}(\alpha, \beta, n)$ is decreasing in group size n. We have

$$\frac{\partial \bar{\sigma}(\alpha, \beta, n)}{\partial n} = \frac{\Delta(\alpha, \beta, n) - [2n(\beta - \alpha) + \beta]\Gamma(\alpha, \beta, n)}{2\Gamma(\alpha, \beta, n)}, \tag{C.3}$$

where $\Gamma(\alpha, \beta, n)$ is given by (19) and where

$$\Delta(\alpha, \beta, n) \equiv \beta^2 n[2n^2 + 3(n-1)] - \alpha\beta[n(4n^2 + 3n - 6) + 1] + 2\alpha^2 n(n^2 - 1). \tag{C.4}$$

The sign of $\partial\bar{\sigma}(\alpha,\beta,n)/\partial n$ is the same as the sign of its numerator. We first determine the sign of $\Delta(\alpha,\beta,n)$. The derivative of $\Delta(\alpha,\beta,n)$ with respect to n is given by

$$\frac{\partial\Delta(\alpha,\beta,n)}{\partial n} = 3\beta^2[2n(n+1)-1] - 6\alpha\beta(2n^2+n-1) + 2\alpha^2(3n^2-1). \tag{C.5}$$

Now, calculating the derivative of this expression with respect to n again, we have

$$\frac{\partial^2\Delta(\alpha,\beta,n)}{\partial n^2} = 6(\beta-\alpha)[\beta+2n(\beta-\alpha)] > 0. \tag{C.6}$$

Hence, $\partial\Delta(\alpha,\beta,n)/\partial n$ is increasing in n and reaches a minimum in $n=2$. We have

$$\frac{\partial\Delta(\alpha,\beta,n)}{\partial n} \geq \frac{\partial\Delta(\alpha,\beta,n)}{\partial n}|_{n=2} = 33\beta^2 - 54\alpha\beta + 22\alpha^2. \tag{C.7}$$

$\partial\Delta(\alpha,\beta,n)/\partial n|_{n=2}$ is convex in β. Thus, $\partial\Delta(\alpha,\beta,n)/\partial n|_{n=2} = 0$ has two solutions in β, given by $\beta \approx 0.77\alpha$ and $\beta \approx 0.87\alpha$. Since $\alpha \in (0,1)$ and $\beta \geq 1$, we can conclude that $\partial\Delta(\alpha,\beta,n)/\partial n|_{n=2}$ is strictly positive implying that $\partial\Delta(\alpha,\beta,n)/\partial n$ is strictly positive for $n \geq 2$. It follows that $\Delta(\alpha,\beta,n)$ is increasing in n and reaches a minimum in $n=2$. In this case, we have

$$\Delta(\alpha,\beta,2) = 22\beta^2 - 33\alpha\beta + 12\alpha^2. \tag{C.8}$$

$\Delta(\alpha,\beta,2)$ is also convex in β. Thus, $\Delta(\alpha,\beta,2) = 0$ has two solutions in β, given by $\beta \approx 0.62\alpha$ and $\beta \approx 0.88\alpha$. Since, again, $\alpha \in (0,1)$ and $\beta \geq 1$, we can conclude that $\Delta(\alpha,\beta,2)$ is strictly positive. It follows that $\Delta(\alpha,\beta,n)$ is strictly positive for any $n \geq 2$.

Recall the sign of $\partial\bar{\sigma}(\alpha,\beta,n)/\partial n$ is the same as the sign of its numerator. Let us define $\Phi(\alpha,\beta,n) = [2n(\beta-\alpha)+\beta]^2\Gamma(\alpha,\beta,n)^2 - \Delta(\alpha,\beta,n)^2$. We have

$$\Phi(\alpha,\beta,n) \equiv 4\beta n(\beta-\alpha)^2[\beta n(2n-3) - 2\alpha(n-1)^2]. \tag{C.9}$$

This term is strictly positive for $\alpha \in (0,1)$ and $\beta \geq 1$ because $\beta n(2n-3) \geq n(2n-3) \geq 2(n-1)^2 > 2\alpha(n-1)^2$ for $n \geq 2$. It follows that $[2n(\beta-\alpha)+\beta]\Gamma(\alpha,\beta,n) > \Delta(\alpha,\beta,n)$, implying that the numerator of $\partial\bar{\sigma}(\alpha,\beta,n)/\partial n$ is negative. Hence, $\bar{\sigma}(\alpha,\beta,n)$ is decreasing in group size n and reaches a maximum in $n=2$. We have

$$\bar{\sigma}(\alpha,\beta,2) = \frac{1}{2}\left[2 + 3\alpha - 6\beta + \sqrt{(10\beta-9\alpha)(2\beta-\alpha)}\right]. \tag{C.10}$$

We have $\bar{\sigma}(\alpha,\beta,2) < 1-\alpha$. Indeed, this inequality is equivalent to $(6\beta-5\alpha)^2 > (10\beta-9\alpha)(2\beta-\alpha)$, which reduces to $16(\beta-\alpha)^2 > 0$. It follows that $\bar{\sigma}(\alpha,\beta,n) < 1-\alpha$ for any $n \geq 2$. Finally, it is easily verified that $\lim_{n\to\infty}\bar{\sigma}(\alpha,\beta,n) = 1-\beta \leq 0$.

D Proof of Proposition 3

First, it is easily verified that $\lim_{\beta\to\infty}\bar{\sigma}(\alpha,\beta,n) = -\infty$. Suppose now that $\beta = 1$ holds. In this case, we have

$$\bar{\sigma}(\alpha,1,n) = \frac{1}{2}\left\{-\left[n^2(1-\alpha)+n+\alpha-2\right] + \Gamma(\alpha,1,n)\right\}. \tag{D.1}$$

where

$$\Gamma(\alpha, 1, n) = \sqrt{(n-1)[n(1-\alpha)+\alpha][n(n+3)-\alpha(n+1)^2]}. \tag{D.2}$$

Let $P(\alpha, n) \equiv \Gamma(\alpha, 1, n)^2 - [(n+1)(n(1-\alpha)+\alpha)-2]^2$. It simplifies to $P(\alpha, n) = 4(1-\alpha)(n-1) > 0$, implying that $\bar{\sigma}(\alpha, 1, n) > 0$ holds for $\alpha \in (0,1)$ and $n \geq 2$. We had that $\partial\bar{\sigma}(\alpha, \beta, n)/\partial\beta$ is negative. It follows that there exists $\beta^*(\alpha, n) > 1$, such that $\bar{\sigma}(\alpha, \beta^*(\alpha, n), n) = 0$. Consequently, if $\beta \in [1, \beta^*(\alpha, n)]$, then $\bar{\sigma}(\alpha, \beta, n) \geq 0$, and thus $K(\alpha, \beta, \sigma, n)$ and $\partial^2 f/\partial\theta'^2 \,|_{\theta'=\theta=\theta^*}$, given respectively by (17) and (16), are both strictly positive (respectively negative) if $\sigma < \bar{\sigma}(\alpha, \beta, n)$ (respectively $\sigma > \bar{\sigma}(\alpha, \beta, n)$). When $\beta > \beta^*(\alpha, n)$, we have $\bar{\sigma}(\alpha, \beta, n) < 0$ and thus $K(\alpha, \beta, \sigma, n)$ and $\partial^2 f/\partial\theta'^2 \,|_{\theta'=\theta=\theta^*}$ are both always strictly negative.

E Linear-quadratic payoffs: detailed analysis

Following Bester and Güth (1998) and Heiftez et al. (2007a), consider a two-player game with material payoff

$$\Pi(x, y) = x(ky + m - x), \tag{E.1}$$

where k captures the effect of the partner's effort on one's own payoff and $m > 0$. The utility of a mutant with action x and preference trait τ is

$$V(x, y) = \Pi(x, y) + \tau\Pi(y, x), \tag{E.2}$$

while the utility of a resident with action y and trait γ is

$$V(y, x) = \Pi(y, x) + \gamma\Pi(x, y). \tag{E.3}$$

These utility functions are strictly concave in own effort, so the game admits a unique interior Nash equilibrium, given by

$$x^*(\tau, \gamma) = \frac{m(2+k(1+\tau))}{4-k^2(1+\tau)(1+\gamma)}, \qquad y^*(\tau, \gamma) = \frac{m(2+k(1+\gamma))}{4-k^2(1+\tau)(1+\gamma)}, \tag{E.4}$$

whenever $4 - k^2(1+\tau)(1+\gamma) \neq 0$.

Let $\Pi^*(\tau, \gamma) \equiv \Pi(x^*(\tau, \gamma), y^*(\tau, \gamma))$ be the equilibrium material payoff of the mutant. The invasion fitness is $f(\tau, \gamma) = \Pi^*(\tau, \gamma) - \Pi^*(\gamma, \gamma)$, leading to the selection gradient

$$D(\gamma) = \left.\frac{\partial f}{\partial \tau}\right|_{\tau=\gamma} = \frac{k^2 m^2[k-\gamma(2-k)]}{[2+k(1+\gamma)][2-k(1+\gamma)]^3}. \tag{E.5}$$

The unique singular point is

$$\gamma^* = \frac{k}{2-k}, \tag{E.6}$$

which is negative when $k < 0$ and positive when $k > 0$.

The derivative of the selection gradient at the singular point is

$$\left.\frac{dD}{d\gamma}\right|_{\gamma=\gamma^*} = -\frac{k^2m^2(2-k)^5}{256(1-k)^3}, \tag{E.7}$$

and the second derivative of mutant fitness with respect to τ at the singular point is

$$\left.\frac{\partial^2 f}{\partial\tau^2}\right|_{\tau=\gamma=\gamma^*} = -\frac{k^2m^2(2+k)(2-k)^5}{512(1-k)^3}. \tag{E.8}$$

For $k \in (-1,0) \cup (0,1)$, both expressions are strictly negative, so the singular point is both convergence stable and evolutionarily stable—hence a continuously stable preference trait.

The singular point is an evolutionary branching point (convergence stable but evolutionarily unstable) if and only if $k < -2$: that is, the negative effect of the partner's effort on one's payoff must exceed twice the own quadratic cost effect.

References

Adams, C. P. (2006). Optimal team incentives with ces production. *Economic Letters*, 92:143–148.

Akçay, E. and Cleve, J. V. (2012). Behavioral responses in structured populations pave the way to group optimality. *The American Naturalist*, 179(2):257–269.

Akçay, E., Van Cleve, J., Feldman, M. W., and Roughgarden, J. (2009). A theory for the evolution of other-regard integrating proximate and ultimate perspectives. *Proceedings of the National Academy of Sciences*, 106(45):19061–19066.

Alger, I. (2023). Evolutionarily stable preferences. *Philosophical Transactions of the Royal Society B*, 378(1876):20210505.

Alger, I. and Weibull, J. (2013). Homo moralis – preference evolution under incomplete information and assortative matching. *Econometrica*, 81:2269–2302.

Alger, I. and Weibull, J. (2019). Evolutionary models of preference formation. *Annual Review of Economics*, 11:329–354.

Alger, I. and Weibull, J. D. (2012). A generalization of hamilton's rule—love others how much? *Journal of Theoretical Biology*, 299:42–54.

Alger, I., Weibull, J. W., and Lehmann, L. (2020). Evolution of preferences in structured populations: Genes, guns, and culture. *Journal of Economic Theory*, 185:104951.

Archetti, M. and Scheuring, I. (2011). Coexistence of cooperation and defection in public goods games. *Evolution*, 65(4):1140–1148.

Bergstrom, T., Blume, L., and Varian, H. (1986). On the private provision of public goods. *Journal of Public Economics*, 29(1):25–49.

Bester, H. and Güth, W. (1998). Is altruism evolutionarily stable? *Journal of Economic Behavior & Organization*, 34(2):193–209.

Bowles, S. and Gintis, H. (2011). *A Cooperative Species: Human Reciprocity and Its Evolution*. Princeton University Press.

Boyd, R. and Richerson, P. J. (1985). *Culture and the evolutionary process*. University of Chicago press.

Boyd, R. and Richerson, P. J. (2005). *The Origin and Evolution of Cultures*. Oxford University Press.

Buchholz, W. and Sandler, T. (2021). Global public goods. *Journal of Economic Literature*, 59:488–545.

Bulow, J. I., D., G. J., and Klempererer, P. D. (1985). Multimarket oligopoly: strategic substitutes and com plements. *Journal of Political Economy*, 93:488–511.

Cavalli-Sforza, L. L. and Feldman, M. W. (1981). *Cultural transmission and evolution: A quantitative approach*. Number 16. Princeton University Press.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics*, 14:47–83.

Cheikbossian, G. (2021a). Evolutionarily stable in-group altruism in intergroup conflict over (local) public goods. *Games and Economic Behavior*, 127:206–226.

Cheikbossian, G. (2021b). The evolutionary stability of in-group altruism in productive and destructive group contests. *Journal of Economic Behavior and Organization*, 188:236–252.

Cheikbossian, G. and Fayat, R. (2018). Group size, collective action and complementarities in efforts. *Economic Letters*, 168:77–81.

Cornes, R. and Hartley, R. (2007). Weak links, good shots and other public good games: Building on bbv. *Journal of Public Economics*, 91:1684–1707.

Dekel, E., C., E. J., and Yilankaya, O. (2007). Evolution of preferences. *Review of Economic Studies*, 74:685–704.

Dieckmann, U. and Law, R. (1996). *The Dynamical Theory of Coevolution: A Derivation from Stochastic Ecological Processes*, volume 34. Journal of Mathematical Biology.

Doebeli, M., Hauert, C., and Killingback, T. (2004). The evolutionary origin of cooperators and defectors. *Science*, 306(5697):859–862.

Eshel, I. (1983). Evolutionary and continuous stability. *Journal of Theoretical Biology*, 103(1):99–111.

Eshel, I. and Motro, U. (1981). Kin selection and strong evolutionary stability of mutual help. *Theoretical Population Biology*, 19(3):420–433.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114(3):817–868.

Geritz, S. A., Kisdi, E., Meszéna, G., and Metz, J. A. (1998). Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evolutionary Ecology*, 12(1):35–57.

Gore, J., Youk, H., and Van Oudenaarden, A. (2009). Snowdrift game dynamics and facultative cheating in yeast. *Nature*, 459(7244):253–256.

Griffin, A. S., West, S. A., and Buckling, A. (2004). Cooperation and competition in pathogenic bacteria. *Nature*, 430(7003):1024–1027.

Güth, W. and Yaari, M. E. (1992). An evolutionary approach to explaining cooperative behavior by reciprocal incentives. *International Journal of Game Theory*, 21(1):151–159.

Heiftez, A., Shannon, C., and Spiegel, Y. (2007a). The dynamic evolution of preferences. *Economic Theory*, 32:251–286.

Heiftez, A., Shannon, C., and Spiegel, Y. (2007b). What to maximize if you must. *Journal of Economic Theory*, 133:31–57.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., et al. (2005). "economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and brain sciences*, 28(6):795–815.

Lehmann, L. and Mullon, C. (2025). Evolution of quantitative traits: exploring the ecological, social and genetic bases of adaptive polymorphism. *bioRxiv*.

McNamara, J. M., Gasson, C. E., and Houston, A. I. (1999). Incorporating rules for responding into evolutionary games. *Nature*, 401(6751):368–371.

Ok, E. A. and Vega-Redondo, F. (2001). On the evolution of individualistic preferences: an incomplete information scenario. *Journal of Economic Theory*, 97:231–254.

Ray, D., Baland, J.-M., and Dagnelie, O. (2007). Inequality and inefficiency in joint projects. *Economic Journal*, 117(522):922–935.

Sethi, R. and Somnathan, R. (2001). Preference evolution and reciprocity. *Journal of Economic Theory*, 97:273–297.

Smith, K. M., Mabulla, I. A., and L., A. C. (2022). Hadza hunter–gatherers with greater exposure to other cultures share more with generous campmates. *Biology Letters*, 18:20220157.

Stibbard-Hawkes, D. N., Smith, K. M., and Apicella, C. L. (2022). Why hunt? why gather? why share? hadza assessments of foraging and food-sharing motive. *Evolution and Human Behavior*, 43:257–272.