

# Intergenerational Cultural Transmission as an Evolutionary Game

James D Montgomery\*  
Department of Sociology  
University of Wisconsin-Madison  
Madison, WI 53706  
jmontgom@ssc.wisc.edu

July 2008

This version: October 23, 2008

**Abstract** We generalize the Bisin and Verdier (2000, 2001) model of intergenerational transmission to permit an arbitrary number of cultural traits. A key observation – that this model is equivalent to an evolutionary game under replicator dynamics – allows us to exploit the extensive literature on replicator systems. For some substantively interesting special cases, we demonstrate global stability of equilibrium outcomes using recent results on stable games and potential games. Drawing on mathematical biology, we also apply results on the long-term survival (permanence) of traits, and highlight the analogy between catalytic networks and systems of cultural traits.

---

\*I am grateful to Bill Sandholm for helpful comments and suggestions, and especially for pointing me toward the mathematical biology literature on permanence.

# 1 Introduction

Sociologists have long studied intergenerational (parent-to-child) transmission of cultural traits such as ethnic identity or religious affiliation.<sup>1</sup> More recently, Bisin and Verdier (2000, 2001) have developed an economic model of intergenerational cultural transmission, incorporating the important insight that parental socialization efforts will vary with the composition of the population. Intuitively, in a society with two traits where each trait has some “cultural distaste” for the other, parents holding the minority trait have greater incentive to incur socialization costs. If these parents are unable to instill their own trait through “direct” socialization, their children are likely to acquire the majority trait through “oblique” socialization (adopting the trait of a “cultural parent” drawn randomly from the population). In contrast, majority parents have less incentive to incur socialization costs because their children are likely to adopt the majority trait through oblique socialization even if direct socialization fails. Moving to the population level, the Bisin-Verdier model predicts long-run persistence of cultural minorities. Parents belonging to a minority group on the verge of “extinction” would choose high socialization effort, thereby increasing the fraction of the population holding the trait. Thus, against the image of a cultural “melting pot” in which all individuals eventually adopt the majority trait, the model suggests that ethnic enclaves or minority religions will persist indefinitely (see Bisin and Verdier 2000 for further discussion).

While their model highlights an important aspect of intergenerational transmission, Bisin and Verdier (2000, 2001) restrict attention to the case of two cultural traits. Recognizing that many societies have more (often many more) than two ethnic or religious groups, the present paper considers the general case with an arbitrary number of traits.<sup>2</sup> Our examination of the  $n$ -trait case is facilitated by a key observation that offers a new perspective on intergenerational transmission and opens new avenues for formal analysis. Namely, given a quadratic specification of the socialization cost function, the Bisin-Verdier model is equivalent to an evolutionary game under replicator dynamics. On this view, “traits” become strategies, the “cultural distaste” of trait  $i$  for trait  $j$  becomes element  $ij$  of the payoff matrix, and stable long-run outcomes can be understood as Nash equilibria.

Beyond simply providing a new game-theoretic perspective, this key observation allows us to exploit the existing literature on replicator systems, drawing on a range of results proven by economic theorists (Sandholm 2008) and mathematical biologists (Hofbauer and Sigmund 1998). For several special cases of the model, corresponding to substantively meaningful restrictions on the payoff matrix, we show that global

---

<sup>1</sup>See Hout and Goldstein (1994) and Hout et al (2001) for recent research and references.

<sup>2</sup>In their empirical work on religious traits, Bisin et al (2004) partition the US population into four groups (Protestant, Catholic, Jew, and other). More recently, Bisin et al (2008) have themselves provided some theoretical analysis of the  $n$ -trait case. We indicate below where the present analysis parallels (and goes beyond) that paper.

stability of Nash equilibria follows immediately from the observation that the game is a stable game (Hofbauer and Sandholm 2008) or a potential game (Sandholm 2001). Further, shifting focus from global stability to the more basic question of survival – whether any trait will become “extinct” in the long run – we review and illustrate some relevant results from the mathematical biology literature on permanence (Schuster et al 1979). Fortuitously, biologists have already studied the permanence of replicator systems under the precise conditions on the payoff matrix imposed by the Bisin-Verdier model (i.e., non-negative cultural distastes). We thus pay special attention to existing results on catalytic networks (Eigen and Schuster 1979) and highlight the analogy between these networks (which arise in the study of prebiotic evolution) and systems of cultural traits.

The paper proceeds as follows. Section 2 generalizes the Bisin-Verdier model to permit  $n \geq 2$  traits, and reveals the equivalence of this model to a replicator system. While this equivalence depends on a particular specification of the socialization cost function, Section 3 shows that the framework is flexible enough to encompass variation across traits in socialization costs and fertility rates. Section 4 examines several special cases of the model, utilizing results on stable and potential games to establish global stability of Nash equilibria. Section 5 offers an analysis of the generic 3-trait case. Section 6 reviews and illustrates results from the mathematical biology literature on permanence. Section 7 contains some concluding remarks.

## 2 Model

Our specification of the model closely follows Bisin and Verdier (2001) and Bisin et al (2008). Consider a large population (formally, a continuum) in which each individual possesses some trait  $i \in \{1, \dots, n\}$ . Let  $q_i$  denote the fraction of the population with trait  $i$ , so that  $\sum_i q_i = 1$ .

Parents reproduce asexually, and children then adopt traits through the following two-stage process.<sup>3</sup> In the first stage, each parent attempts to instill her own trait through “direct” socialization. Given the parent’s trait  $i$ , direct socialization succeeds with probability  $d_i$ . If direct socialization fails (probability  $1 - d_i$ ), the child’s trait is determined in the second stage through “oblique” socialization. That is, the child adopts the trait of a “cultural parent” drawn randomly from the population.<sup>4</sup> Thus, a parent with trait  $i$  will have a child with trait  $j \neq i$  with probability

$$P_{ij} = (1 - d_i)q_j, \tag{1}$$

and will have a child with trait  $i$  with probability

$$P_{ii} = d_i + (1 - d_i)q_i. \tag{2}$$

---

<sup>3</sup>See Bisin and Verdier (2000) for an extension of the model in which reproduction is sexual, and the model thus addresses matching in the marriage market.

<sup>4</sup>My use of terminology (“direct” and “oblique” and “cultural parent”) follows Bisin and Verdier (2000, 2001), who in turn borrowed it from Cavalli-Sforza and Feldman (1981).

It is important to note that, even if direct socialization fails, the child will still adopt the parent's trait if the cultural parent happens to possess that trait.

Although our analysis will focus on continuous-time population dynamics, it may be helpful to begin with a simple overlapping-generations specification (following Bisin and Verdier 2000). If each individual lives one period (generation  $t$ ) and then has one child (generation  $t + 1$ ), population dynamics are determined by the system of  $n$  difference equations

$$q_i(t + 1) = \sum_j P_{ji} q_j(t) \quad \forall i. \quad (3)$$

Substituting (1) and (2) into (3) yields

$$q_i(t + 1) = q_i(t) + q_i(t) [d_i - \sum_j d_j q_j(t)] \quad \forall i. \quad (4)$$

Moving to continuous time, we thus obtain

$$\dot{q}_i = q_i [d_i - \sum_j d_j q_j] \quad \forall i. \quad (5)$$

This system of differential equations may be viewed simply as a continuous-time approximation to (4). Alternatively, as shown in the next section, (5) may also be derived as the limiting case given a slightly different (non-overlapping generations) specification of population dynamics.

Regardless of the precise specification of the dynamics (discrete or continuous), one key observation made by Bisin and Verdier (2000, 2001) is already apparent. Inspection of (4) or (5) reveals that the proportion of the population with trait  $i$  is rising when the probability of direct socialization by parents with this trait ( $d_i$ ) exceeds the expected probability of direct socialization in the population ( $\sum_j d_j q_j$ ). Consequently, if direct socialization probabilities are exogenously given, the trait  $i$  with the largest  $d_i$  will continue to grow until the entire population has adopted this trait. Because this long-run homogeneity result is difficult to reconcile with the observed persistence of cultural diversity (e.g., ethnic enclaves or minority religions), Bisin and Verdier (2000, 2001) suggest that the direct socialization probabilities are endogenously determined by parents, and specify the utility maximization problem so that the parent's optimal effort is higher when her trait is less common.

More precisely, suppose that each parent with trait  $i$  chooses the direct socialization probability  $d_i$  to solve

$$\max_{d_i} \sum_j P_{ij} V_{ij} - H(d_i) \quad (6)$$

where  $V_{ij}$  is the value that a parent with trait  $i$  places on a child with trait  $j$ , and  $H(d)$  is a convex cost function. Letting

$$\Delta_{ij} = V_{ii} - V_{ij} \quad (7)$$

denote the “cultural distaste” of trait  $i$  for trait  $j$ , we immediately obtain  $\Delta_{ii} = 0$  for all  $i$ , and further assume

$$\Delta_{ij} \geq 0 \quad \forall i, j \quad (8)$$

to capture the parent’s desire for the child to adopt the parent’s own type.<sup>5</sup> The optimal solution to (6) is determined implicitly by the first-order condition

$$H'(d_i^*) = \sum_j \Delta_{ij} q_j \quad (9)$$

where the right-hand side may be interpreted as the parent’s “expected distaste” of her child’s cultural parent (if direct socialization fails). Convexity of the cost function  $H(d)$  implies that the parent’s optimal choice  $d_i^*$  is an increasing function of this expectation.

Recognizing that the choice variable  $d$  is a probability, we might require the cost function  $H(d)$  to satisfy the condition  $H'(d) \rightarrow \infty$  as  $d \rightarrow 1$ . However, for purposes of tractability, we (following Bisin et al 2008) adopt the functional form

$$H(d) = \frac{1}{2}d^2, \quad (10)$$

and further impose the conditions

$$\Delta_{ij} \leq 1 \quad \forall i, j \quad (11)$$

to ensure an interior solution to the parent’s optimization problem.<sup>6</sup> Given this specification of the cost function, (9) becomes

$$d_i^* = \sum_j \Delta_{ij} q_j \quad (12)$$

so that the parent’s optimal choice is simply a linear function of their expected distaste of cultural parents.

Having endogenized the direct socialization probabilities, we now return to the population dynamics. Substitution of (12) into (5) yields

$$\dot{q}_i = q_i \left[ \sum_j \Delta_{ij} q_j - \sum_j q_j \sum_k \Delta_{jk} q_k \right] \quad \forall i. \quad (13)$$

Adopting matrix notation, this becomes

$$\dot{q}_i = q_i [(\Delta \mathbf{q})_i - \mathbf{q}' \Delta \mathbf{q}] \quad \forall i \quad (14)$$

---

<sup>5</sup>While the existence of non-negative cultural distastes might simply be taken as a primitive assumption, Bisin and Verdier (2001) suggest that these distastes arise from “imperfect empathy” whereby a child’s choice is assessed using the parent’s utility function.

<sup>6</sup>Anticipating the generalization of the model in the next section, we might ensure an interior solution by instead requiring each cost parameter  $c_i$  to be sufficiently large.

where

$$\mathbf{q} = \begin{bmatrix} q_1 \\ \vdots \\ q_n \end{bmatrix} \quad (15)$$

is an  $(n \times 1)$  column vector characterizing the distribution of traits,

$$\Delta = \begin{bmatrix} \Delta_{11} & \cdots & \Delta_{1n} \\ \vdots & \ddots & \vdots \\ \Delta_{n1} & \cdots & \Delta_{nn} \end{bmatrix} \quad (16)$$

is an  $(n \times n)$  matrix characterizing cultural distastes,  $(\Delta \mathbf{q})_i$  denotes the  $i$ th element of the column vector  $\Delta \mathbf{q}$ , and  $\mathbf{q}'$  denotes the transpose of  $\mathbf{q}$ . Equation (14) may be written even more compactly as

$$\dot{\mathbf{q}} = \text{diag}(\mathbf{q})[\Delta \mathbf{q} - \mathbf{q}'\Delta \mathbf{q}] \quad (17)$$

where  $\text{diag}(\mathbf{q})$  is a square matrix with the elements of  $\mathbf{q}$  along the main diagonal and zeros elsewhere.

Evolutionary game theorists will immediately recognize (17) as the equation for the replicator system.<sup>7</sup> Indeed, this simple observation (not made explicitly by Bisin et al) lies at the heart of the present paper. Reinterpreting the Bisin-Verdier model as a population game under replicator dynamics, “traits” become strategies,  $\Delta$  becomes a payoff matrix ( $\Delta_{ij}$  is the payoff received by strategy  $i$  when matched against strategy  $j$ ),  $\Delta \mathbf{q}$  represents the vector of expected payoffs (given random matching), and  $\mathbf{q}'\Delta \mathbf{q}$  represents the average payoff for the population.<sup>8</sup> Given the specification of the replicator dynamics, trait  $i$  grows when the expected payoff for this trait  $(\Delta \mathbf{q})_i$  exceeds the population average payoff  $\mathbf{q}'\Delta \mathbf{q}$ . Adopting this game-theoretic perspective, the key assumption of the Bisin-Verdier model emerges as a restriction on the payoff matrix: each trait receives zero payoff when matched against itself ( $\Delta_{ii} = 0 \forall i$ ) and receives a non-negative payoff when matched against other traits ( $\Delta_{ij} \geq 0 \forall j \neq i$ ).

This game-theoretic reinterpretation of the model is both conceptually interesting and analytically useful. Conceptually, it suggests that we can study intergenerational transmission without explicit attention to socialization processes. On this view, individuals directly choose their own traits to maximize their own expected payoffs,

---

<sup>7</sup>This framework was first introduced by Taylor and Jonker (1978); see Hofbauer and Sigmund (1998) or Sandholm (2008) for textbook treatments.

<sup>8</sup>More generally, a population game is characterized by an  $(n \times 1)$  payoff vector  $\mathbf{F}(\mathbf{q}) = [F_1(\mathbf{q}) \dots F_n(\mathbf{q})]'$ . In the present context, the population game is linear so that  $\mathbf{F}(\mathbf{q}) = \Delta \mathbf{q}$ . Following the standard interpretation in evolutionary game theory, payoffs result from pairwise random matching within a single population given a two-player normal-form game that is symmetric (in the sense that there is no distinction between the row and column players). See Sandholm (2008, Chap 1) for further discussion of population games.

and the expected payoff for each trait is given by its expected distaste of a trait drawn randomly from the population.<sup>9</sup> We thus obtain an interesting new perspective on intergenerational transmission by viewing cultural distastes as the payoffs to traits. Analytically, this reinterpretation of the model allows us to draw upon existing results concerning global stability and permanence in population games. While Bisin et al (2008) have already analyzed one special case of the  $n$ -trait model, we show below that their global stability result follows immediately from the observation that, in this special case, the game is stable (in the sense defined by Hofbauer and Sandholm 2008). Moving beyond Bisin et al (2008), we further show how analysis of several other special cases is facilitated by existing results on classes of population games. Finally, shifting focus from global stability to the long-run survival of traits, we can make use of results from the mathematical biology literature on permanence (Schuster et al 1979), especially in catalytic networks (Eigen and Schuster 1979).

### 3 Generalizing the model

Before moving to the analysis, it is useful to consider how far the Bisin-Verdier model can be generalized without breaking the tight link between this model and the replicator dynamics. Reflection suggests that the linearity of the parent’s optimal choice in their expected distaste of cultural parents (see equation 12) is crucial. While this severely restricts the functional form of the cost function, it is trivial to introduce heterogeneity by respecifying (10) as

$$H_i(d_i) = \frac{c_i}{2} d_i^2 \quad (18)$$

so that the optimal solution for parents with trait  $i$  becomes

$$d_i^* = \sum_j \frac{\Delta_{ij}}{c_i} q_j. \quad (19)$$

Given heterogeneity of this form, (17) becomes

$$\dot{\mathbf{q}} = \text{diag}(\mathbf{q})[\hat{\Delta}\mathbf{q} - \mathbf{q}'\hat{\Delta}\mathbf{q}] \quad (20)$$

where

$$\hat{\Delta} = \begin{bmatrix} \frac{\Delta_{11}}{c_1} & \dots & \frac{\Delta_{1n}}{c_1} \\ \vdots & \ddots & \vdots \\ \frac{\Delta_{n1}}{c_n} & \dots & \frac{\Delta_{nn}}{c_n} \end{bmatrix}. \quad (21)$$

Thus, a change in the socialization cost for trait  $i$  essentially rescales the cultural distastes of this trait (characterized by row  $i$  of the  $\Delta$  matrix) without otherwise affecting the population dynamics.

---

<sup>9</sup>Of course, as we will see in the analysis below, the growth of even the most “intolerant” trait is self-limiting given that traits receive zero payoff when matched against themselves.

Perhaps more interestingly, the model can also accommodate differential fertility (i.e., differences in exogenous reproduction rates) across traits.<sup>10</sup> Let  $r_i$  denote the reproduction rate (i.e., the expected number of children) for a parent with trait  $i$ . To incorporate variation in the number of children into parent's utility function in the simplest possible way, suppose that both costs and benefits are scaled by the realized number of children, so that the parent's optimal choice (19) does not depend on the number of children.<sup>11</sup> Moving to the population dynamics, it is convenient to replace the overlapping generations specification above with a modified version of the non-overlapping generations specification suggested by Bisin et al (2008). Suppose that period length is  $h$ , that a fraction  $h$  of the population dies in the period between time  $t$  and  $t + h$ , and that fertility and socialization of children occur in the last period of life. Population dynamics are then given by

$$q_i(t+h) = \frac{q_i(t)(1-h) + h \sum_j r_j P_{ji} q_j(t)}{1-h + h \sum_j r_j q_j(t)}. \quad (22)$$

Substituting (1) and (2) into (22), some algebra yields

$$\frac{q_i(t+h) - q_i(t)}{h} = \frac{q_i(t)[r_i d_i - \sum_j r_j d_j q_j(t)]}{1-h + h \sum_j r_j q_j(t)} \quad (23)$$

and thus, in the limit as  $h \rightarrow 0$ , we obtain

$$\dot{q}_i = q_i[r_i d_i - \sum_j r_j d_j q_j]. \quad (24)$$

Note that, if each parent has one child ( $r_i = 1$  for all  $i$ ), equation (23) reduces to (4) when  $h = 1$ , while (24) reduces to (5), and thus we have provided a stronger justification of the continuous-time specification.<sup>12</sup> Further substituting the parent's optimal decision (19) into (24), population dynamics are given by

$$\dot{\mathbf{q}} = \text{diag}(\mathbf{q})[\tilde{\Delta}\mathbf{q} - \mathbf{q}'\tilde{\Delta}\mathbf{q}] \quad (25)$$

where

$$\tilde{\Delta} = \begin{bmatrix} \frac{r_1 \Delta_{11}}{c_1} & \dots & \frac{r_1 \Delta_{1n}}{c_1} \\ \vdots & \ddots & \vdots \\ \frac{r_n \Delta_{n1}}{c_n} & \dots & \frac{r_n \Delta_{nn}}{c_n} \end{bmatrix}. \quad (26)$$

<sup>10</sup>See Bisin and Verdier (2001) for discussion of endogenous fertility.

<sup>11</sup>We could further introduce economies of scale in socialization, specifying the parent's cost function as  $g_i(k)d_i^2$  where  $k$  is the realized number of children and  $g_i(k)$  is a concave function. However, given this non-linearity, the optimal choice  $d_i^*$  would depend on  $k$ , and the average  $d_i^*$  for parents with trait  $i$  would depend on the probability distribution for  $k$  (not merely its mean,  $r_i$ ). For simplicity, we thus retain the linear specification  $g_i(k) = \frac{c_i}{2}k$  which yields (19).

<sup>12</sup>More generally, if  $r_i = r$  for all  $i$ , we obtain  $\dot{q}_i = r q_i [d_i - \sum_j d_j q_j]$ , and observe that a change in the (common) reproduction rate merely affects the "velocity" of the system without affecting its trajectories (see Hofbauer and Sigmund 1998, p 32).

Thus, differential fertility also rescales the rows of the  $\Delta$  matrix without otherwise affecting the population dynamics.

For simplicity, we revert to equation (17) for the remainder of this paper. Nevertheless, readers might keep in mind that cultural distastes characterized by the  $\Delta$  matrix may reflect not only “pure” distastes but also heterogeneity in socialization costs and reproduction rates.

## 4 Analysis of some special cases

Moving now to the analysis, we consider several special cases obtained by imposing substantively meaningful restrictions on the payoff matrix. For each case, we show how global stability of Nash equilibria can be easily established using recent results from evolutionary game theory. To streamline the analysis, we adopt (in both this section and the next) the simplifying assumption that all cultural distastes are strictly positive ( $\Delta_{ij} > 0 \forall j \neq i$ ). Recognizing that the  $\Delta_{ij}$ 's may be arbitrarily close to zero, this assumption has essentially no substantive implication, but simplifies statements concerning the dynamics and equilibrium of the model.<sup>13</sup>

### 4.1 Stable games

In their analysis of the  $n$ -trait model, Bisin et al (2008) maintain two strong assumptions. First, they adopt the specification of the cost function in (10), which we also retain throughout this paper. Second, they assume that the cultural distastes satisfy the condition

$$\Delta_{ij} = \Delta_i \quad \forall j \neq i, \quad (27)$$

so that each trait  $i$  is “equally intolerant” of every other trait  $j \neq i$ . Under this condition, we may (without loss of generality) index traits so that

$$\Delta_1 \geq \Delta_2 \geq \dots \geq \Delta_n. \quad (28)$$

That is, trait 1 is the most intolerant trait (with the greatest cultural distaste for other traits), while trait  $n$  is the least intolerant trait.

While Bisin et al (2008) do not recognize it as a Nash equilibrium, they do provide their own characterization of the following equilibrium.

**Proposition 1** *Given (27), the population game  $\Delta \mathbf{q}$  has a unique Nash equilibrium. In this equilibrium, at least two traits are active (i.e., played by a positive fraction of the population). Defining the function*

$$\theta_i(\pi) = \min\{0, 1 - \frac{\pi}{\Delta_i}\}, \quad (29)$$

---

<sup>13</sup>This assumption allows us to ignore some types of outcomes (e.g., a continuum of rest points) that are structurally unstable (i.e., not robust to an arbitrarily small respecification of the dynamical system) and hence non-generic.

this equilibrium is characterized by the system of  $(n + 1)$  equations

$$q_i^* = \theta_i(\pi^*) \quad \forall i \quad (30)$$

$$\sum_i \theta_i(\pi^*) = 1 \quad (31)$$

where  $q_i^*$  is the fraction of the population choosing trait  $i$ , and  $\pi^* \in (0, \Delta_2)$  is the expected payoff of active traits.

**Proof** Given (27), the expected payoff to trait  $i$  is  $(1 - q_i)\Delta_i$ . In equilibrium, all active traits generate the same expected payoff  $\pi^*$ , and any inactive traits would generate an expected payoff below  $\pi^*$ . These conditions are equivalent to (30) because

$$q_i^* > 0 \iff q_i^* = 1 - \frac{\pi^*}{\Delta_i} \iff (1 - q_i^*)\Delta_i = \pi^* \quad (32)$$

and

$$q_i^* = 0 \iff 1 - \frac{\pi^*}{\Delta_i} < 0 \iff \Delta_i < \pi^*. \quad (33)$$

To establish the uniqueness of this equilibrium, note that  $\theta_i(0) = 1$  and that  $\theta_i(\pi)$  is weakly decreasing in  $\pi$  for all  $i$ . Consequently,  $\sum_i \theta_i(0) = n$  and  $\sum_i \theta_i(\pi)$  is weakly decreasing in  $\pi$ . Given  $\pi = \Delta_2$ , we obtain  $\theta_i(\Delta_2) = 0$  for all  $i \in \{2, \dots, n\}$  while  $\theta_1(\Delta_2) = 1 - \frac{\Delta_2}{\Delta_1}$ . Thus,  $\sum_i \theta_i(\Delta_2) = 1 - \frac{\Delta_2}{\Delta_1}$ , which is less than 1. Consequently, there must be a unique value  $\pi^* \in (0, \Delta_2)$  such that condition (31) holds. Finally, because  $\Delta_1 \geq \Delta_2 > \pi^*$ , we obtain  $q_1^* \geq q_2^* > 0$ . That is, at least two traits are active in equilibrium.  $\square$

Thus, in the unique Nash equilibrium, the fraction of the population adopting trait  $i$  is directly related to trait  $i$ 's distaste for other traits. The  $m$  most intolerant traits are active (for some  $m \in \{2, \dots, n\}$ ), while the remaining traits are inactive, being “too tolerant” to remain viable.

While any Nash equilibrium is a rest point of the replicator dynamics, there may be many rest points beyond the one identified in Proposition 1. In particular, given the  $m$  traits that are active in the Nash equilibrium, there is a rest point of the replicator dynamics with the set of active traits  $M$  for every non-empty subset  $M \subseteq \{1, \dots, m\}$ . Intuitively, any trait that is “zeroed out” will never be revived by the replicator dynamics. Although we could attempt to directly assess the local stability of these rest points on the boundary of the  $n$ -dimensional simplex, it is easy to see that they cannot be stable. If one of these rest points was stable, then it must be a Nash equilibrium (see Samuelson 2002, p 58), but this would contradict the uniqueness result from Proposition 1.

Still, we have not yet established local (or global) stability of the Nash equilibrium. In their analysis of the present case, Bisin et al (2008) prove that the

equilibrium described in Proposition 1 is locally stable, and then offer simulation analysis which suggests global stability. That is, for every initial condition  $\mathbf{q}(0)$  in the interior of the  $n$ -dimensional simplex, their simulation analysis suggests that the trajectory  $\{\mathbf{q}(t)\}_{t \geq 0}$  will converge to the Nash equilibrium as  $t \rightarrow \infty$ .

Condition (27) does, in fact, imply global stability. Indeed, having recast the Bisin-Verdier model as an evolutionary game, it is simple to prove this result (without recourse to simulation analysis) by applying a recent result from that literature. Following Hofbauer and Sandholm (2008), a population game is *stable* when it exhibits “self-defeating externalities.” Formally, a linear population game with payoff matrix  $\Delta$  is stable when

$$(\mathbf{y} - \mathbf{x})'(\Delta\mathbf{y} - \Delta\mathbf{x}) \leq 0 \quad \forall \mathbf{x}, \mathbf{y} \quad (34)$$

where  $\mathbf{x}$  and  $\mathbf{y}$  are probability vectors (i.e., distributions). To interpret this condition, note that a change in the distribution from  $\mathbf{x}$  to  $\mathbf{y}$  causes the expected payoff for strategy  $i$  to change by  $(\Delta\mathbf{y} - \Delta\mathbf{x})_i$ . Multiplying by the change in the fraction of population playing this strategy, we obtain  $(\mathbf{y} - \mathbf{x})_i(\Delta\mathbf{y} - \Delta\mathbf{x})_i$ . Summing over all strategies, the game is stable when this sum is not positive (and is *strictly stable* when this sum is negative). Condition (34) may be written more compactly as

$$\mathbf{z}'\Delta\mathbf{z} \leq 0 \quad \forall \mathbf{z} \quad (35)$$

where  $\mathbf{z}$  is a displacement vector (satisfying  $\sum_i z_i = 0$ ). Thus, a linear game is stable when its payoff matrix  $\Delta$  is negative semi-definite with respect to all displacement vectors. Stable games are of interest in the present context because of

**Theorem 1 (Sandholm 2008)** *Consider a population game that is strictly stable and has a unique Nash equilibrium. Under the replicator dynamics, every trajectory in the interior of the  $n$ -dimensional simplex converges to the Nash equilibrium.*

See Sandholm (2008, Theorem 6.2.4) for a more precise statement and proof.

Returning to the present model, it is straightforward to verify that the game is stable under condition (27). Applying condition (35) with an arbitrary displacement vector  $\mathbf{z}$ , we obtain

$$\Delta\mathbf{z} = \begin{bmatrix} 0 & \Delta_1 & \dots & \Delta_1 & \Delta_1 \\ \Delta_2 & 0 & \dots & \Delta_2 & \Delta_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Delta_{n-1} & \Delta_{n-1} & \dots & 0 & \Delta_{n-1} \\ \Delta_n & \Delta_n & \dots & \Delta_n & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_{n-1} \\ -\sum_{i \neq n} z_i \end{bmatrix} = \begin{bmatrix} -\Delta_1 z_1 \\ -\Delta_2 z_2 \\ \vdots \\ -\Delta_{n-1} z_{n-1} \\ \Delta_n \sum_{i \neq n} z_i \end{bmatrix} \quad (36)$$

and hence

$$\mathbf{z}'\Delta\mathbf{z} = -\sum_{i \neq n} \Delta_i z_i^2 - \Delta_n \left(\sum_{i \neq n} z_i\right)^2 = -\sum_i \Delta_i z_i^2 < 0 \quad (37)$$

for all displacement vectors  $\mathbf{z}$ . Having already established the existence of a unique Nash equilibrium in Proposition 1, Sandholm's result thus guarantees global stability of this equilibrium.

We have so far restricted attention to the special case of the Bisin-Verdier model in which the payoff matrix satisfies condition (27), though I have allowed an arbitrary number of traits. In contrast, Bisin and Verdier (2001) consider the general case (with no restriction on the payoff matrix) where  $n = 2$ . But given that the payoff matrix

$$\Delta = \begin{bmatrix} 0 & \Delta_1 \\ \Delta_2 & 0 \end{bmatrix} \quad (38)$$

is generic for the 2-trait case, results for the general 2-trait case are a simple corollary to the results just established. That is, there is a unique Nash equilibrium where

$$q_1^* = \frac{\Delta_1}{\Delta_1 + \Delta_2}, \quad q_2^* = \frac{\Delta_2}{\Delta_1 + \Delta_2}, \quad \text{and } \pi^* = \frac{\Delta_1 \Delta_2}{\Delta_1 + \Delta_2}, \quad (39)$$

and this equilibrium is globally stable.<sup>14</sup>

Moving beyond Bisin et al (2008), analysis of another special case also follows immediately. Given any matrix  $\Delta$  that satisfies (35), its transpose  $\Delta'$  will also satisfy this condition.<sup>15</sup> Thus, if the population game  $\Delta' \mathbf{q}$  has a unique Nash equilibrium, the global stability of this equilibrium is already established. In particular, we may transpose (27) to obtain the condition

$$\Delta_{ij} = \Delta_j \quad \forall i \neq j \quad (40)$$

so that we now assume common payoffs in each column (not row) of the payoff matrix. Substantively, every trait  $i \neq j$  has the same distaste for trait  $j$ .

**Proposition 2** *Given (40), the population game  $\Delta \mathbf{q}$  has a unique Nash equilibrium. In this equilibrium, all traits are active, and the fraction of the population choosing trait  $i$  is given by*

$$q_i^* = \frac{1}{\Delta_i \sum_j (1/\Delta_j)}. \quad (41)$$

**Proof** Given (40), the expected payoff for trait  $i$  is given by

$$\sum_{j \neq i} \Delta_j q_j = -\Delta_i q_i + \sum_j \Delta_j q_j. \quad (42)$$

In equilibrium, all active traits generate the same expected payoff  $\pi^*$ , and any inactive trait generates an expected payoff below  $\pi^*$ . Assuming that all traits are active, we thus obtain

$$q_i^* = \frac{\nu^*}{\Delta_i} \quad \forall i \quad (43)$$

---

<sup>14</sup>Of course, this result also presumes the cost function (10), while Bisin and Verider (2001) permit more general cost functions.

<sup>15</sup>We simply transpose both sides of (35) to obtain  $(\mathbf{z}' \Delta \mathbf{z})' = \mathbf{z}' \Delta' \mathbf{z} \leq 0' = 0$ .

where

$$\nu^* = -\pi^* + \sum_j \Delta_j q_j^*. \quad (44)$$

Given that the  $q_i^*$ 's must sum to unity, we obtain

$$\nu^* = \frac{1}{\sum_j (1/\Delta_j)}, \quad (45)$$

and substitution into (43) yields (41). To show that all traits are active in equilibrium (and thus establish uniqueness), note that any inactive trait  $i$  (with  $q_i^* = 0$ ) would generate expected payoff

$$\sum_j \Delta_j q_j^* > \pi^*. \quad (46)$$

Because any inactive trait would generate a higher expected payoff than active traits, all traits must be active in equilibrium.  $\square$

Thus, in the unique Nash equilibrium, all traits are active, and the fraction of the population adopting trait  $i$  is inversely related to the distaste of other traits for  $i$ . As already discussed, global stability of this equilibrium is guaranteed by (37).

## 4.2 Potential games

Moving to another special case, suppose that the payoff matrix is symmetric, so that

$$\Delta_{ij} = \Delta_{ji} \quad \forall i, j. \quad (47)$$

Substantively, trait  $i$ 's distaste for trait  $j$  is equal to  $j$ 's distaste for  $i$  for all pairs of traits  $\{i, j\}$ . Unlike the special cases considered above, multiple Nash equilibria are now possible. For instance, given

$$\Delta = \begin{bmatrix} 0 & 1 & \epsilon & \epsilon \\ 1 & 0 & \epsilon & \epsilon \\ \epsilon & \epsilon & 0 & 1 \\ \epsilon & \epsilon & 1 & 0 \end{bmatrix}, \quad (48)$$

where  $\epsilon$  is small, it is easy to see that the Nash equilibria are

$$\mathbf{q}^* = \begin{bmatrix} 1/2 \\ 1/2 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{q}^{**} = \begin{bmatrix} 0 \\ 0 \\ 1/2 \\ 1/2 \end{bmatrix}, \quad \text{and} \quad \mathbf{q}^{***} = \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}. \quad (49)$$

Focusing on this example, we might proceed to directly verify local stability of  $\mathbf{q}^*$  and  $\mathbf{q}^{**}$  as well as instability of  $\mathbf{q}^{***}$ . But analysis is again facilitated by another recent result from evolutionary game theory.

Following Sandholm (2001), a population game  $\Delta \mathbf{q}$  is a *potential* game when all relevant information about payoffs can be summarized by a single-valued (potential) function  $f(\mathbf{q})$ . More precisely, the potential function must satisfy

$$\frac{\partial f(\mathbf{q})}{\partial q_i} = (\Delta \mathbf{q})_i \quad \forall i. \quad (50)$$

It can be shown that a potential function exists if and only if  $\Delta$  is symmetric, and in this case, the potential function is simply

$$f(\mathbf{q}) = \frac{1}{2} \mathbf{q}' \Delta \mathbf{q}. \quad (51)$$

The existence of a potential function greatly simplifies analysis of global stability. Any local maximum of the potential function is a Nash equilibrium. Further, under the replicator dynamics, potential rises along every trajectory. Together, these results imply

**Theorem 2 (Sandholm 2008)** *Given a potential game under replicator dynamics, every trajectory converges to some Nash equilibrium.*

See again Sandholm (2008, Theorem 6.1.3) for a more precise statement and proof. At first glance, this result might seem rather weak, since it is well known that any stable rest point of the replicator dynamics is a Nash equilibrium (Samuelson 2002). But note that Theorem 2 guarantees local stability of at least one Nash equilibrium, and also rules out the possibility of more complicated (cyclical or chaotic) attractors.

Returning to the example in equation (48), the potential function is

$$f(\mathbf{q}) = q_1 q_2 + q_3 q_4 + \epsilon (q_1 + q_2)(q_3 + q_4), \quad (52)$$

and it is evident that (for small  $\epsilon$ ) both  $\mathbf{q}^*$  and  $\mathbf{q}^{**}$  are local maxima of the potential function, while  $\mathbf{q}^{***}$  is a local minimum. Consequently, every trajectory must converge either to  $\mathbf{q}^*$  or  $\mathbf{q}^{**}$ .

While a potential function may have more than one local maximum, a unique maximum is guaranteed when the potential function is concave. In this case, the payoff matrix must  $\Delta$  satisfy not only symmetry (47) but also stability (35). In particular, for  $n = 3$ , the stability condition becomes

$$\Delta_{12} z_1 z_2 + \Delta_{13} z_1 z_3 + \Delta_{23} z_2 z_3 \leq 0, \quad (53)$$

and the potential function is concave under the conditions

$$\begin{aligned} \psi_1 &= \Delta_{12} + \Delta_{13} - \Delta_{23} > 0 \\ \psi_2 &= \Delta_{12} + \Delta_{23} - \Delta_{13} > 0 \\ \psi_3 &= \Delta_{13} + \Delta_{23} - \Delta_{12} > 0. \end{aligned} \quad (54)$$

To see this, note that the equation

$$(\psi_1 z_1 + \psi_2 z_2 + \psi_3 z_3)(z_1 + z_2 + z_3) = 0 \quad (55)$$

must hold for any displacement vector  $\mathbf{z} = [z_1 \ z_2 \ z_3]'$ . Expanding and rearranging, we obtain

$$\psi_1 z_1^2 + \psi_2 z_2^2 + \psi_3 z_3^2 = -2[\Delta_{12} z_1 z_2 + \Delta_{13} z_1 z_3 + \Delta_{23} z_2 z_3], \quad (56)$$

and thus (53) holds because the left-hand side of (56) must be positive. Loosely, condition (54) requires that the cultural distastes  $\Delta_{12}$ ,  $\Delta_{13}$ , and  $\Delta_{23}$  not be “too far” apart. But anticipating our more general analysis of the  $n = 3$  case in the next section, we can also interpret (54) as the “invasion conditions” (see equation 58 below) in the special case where  $\Delta$  is symmetric.

## 5 The general three-trait case

In this section, we consider the three-trait case where the payoff matrix is generically

$$\Delta = \begin{bmatrix} 0 & \Delta_{12} & \Delta_{13} \\ \Delta_{21} & 0 & \Delta_{23} \\ \Delta_{31} & \Delta_{32} & 0 \end{bmatrix}, \quad (57)$$

and we continue to assume that all cultural distastes are strictly positive. In this case, the outcome hinges on the signs of the following expressions:

$$\begin{aligned} \Psi_1 &= \Delta_{12}\Delta_{23} + \Delta_{13}\Delta_{32} - \Delta_{23}\Delta_{32} \\ \Psi_2 &= \Delta_{21}\Delta_{13} + \Delta_{23}\Delta_{31} - \Delta_{13}\Delta_{31} \\ \Psi_3 &= \Delta_{31}\Delta_{12} + \Delta_{32}\Delta_{21} - \Delta_{12}\Delta_{21} \end{aligned} \quad (58)$$

To motivate these expressions, suppose that trait 3 has been “zeroed out” ( $q_3 = 0$ ) and the population game has converged to the unique Nash equilibrium in which traits 1 and 2 are active. Drawing upon our analysis of the two-trait case above (cf equation 39), this equilibrium entails

$$q_1^* = \frac{\Delta_{12}}{\Delta_{12} + \Delta_{21}}, \quad q_2^* = \frac{\Delta_{21}}{\Delta_{12} + \Delta_{21}}, \quad \text{and} \quad \pi^* = \frac{\Delta_{12}\Delta_{21}}{\Delta_{12} + \Delta_{21}}. \quad (59)$$

This equilibrium always constitutes a rest point of the replicator dynamics. However, trait 3 can successfully “invade” if its expected payoff exceeds the population average payoff, so that

$$\Delta_{31} q_1^* + \Delta_{32} q_2^* > \pi^*. \quad (60)$$

Substituting (59) into (60), we obtain  $\Psi_3 > 0$ , which may thus be interpreted as the “invasion condition” for trait 3. We obtain the other invasion conditions  $\Psi_1 > 0$

(under which trait 1 can invade traits 2 and 3) and  $\Psi_2 > 0$  (under which 2 can invade 1 and 3) analogously.

To characterize the Nash equilibrium, it is first useful to establish the following

**Lemma** *The invasion conditions must hold for at least two traits.*

**Proof** The definitions (58) imply

$$\begin{aligned}\Delta_{12}\Psi_2 + \Delta_{13}\Psi_3 &= \pi \\ \Delta_{21}\Psi_1 + \Delta_{23}\Psi_3 &= \pi \\ \Delta_{31}\Psi_1 + \Delta_{32}\Psi_2 &= \pi\end{aligned}\tag{61}$$

where

$$\pi = \Delta_{12}\Delta_{23}\Delta_{31} + \Delta_{13}\Delta_{32}\Delta_{21} > 0.\tag{62}$$

If any pair of invasion conditions did not hold, the left-hand side of one of the equations in (61) would be non-positive, contradicting the inequality (62).  $\square$

Consequently, we may restrict attention to two cases, corresponding to the number of invasion conditions satisfied given the payoff matrix  $\Delta$ .

**Proposition 3** *For the general case where  $n = 3$ , there are two subcases:*

(1) *Suppose all three invasion conditions hold. There is a unique Nash equilibrium in which*

$$q_1^* = \frac{\Psi_1}{\Psi_1 + \Psi_2 + \Psi_3}, \quad q_2^* = \frac{\Psi_2}{\Psi_1 + \Psi_2 + \Psi_3}, \quad \text{and} \quad q_3^* = \frac{\Psi_3}{\Psi_1 + \Psi_2 + \Psi_3},\tag{63}$$

*and the expected payoff to each trait is given by*

$$\pi^* = \Delta_{12}\Delta_{23}\Delta_{31} + \Delta_{13}\Delta_{32}\Delta_{21}.\tag{64}$$

(2) *Suppose only two invasion conditions hold. Without loss of generality, assume  $\Psi_1 > 0, \Psi_2 > 0$ , and  $\Psi_3 \leq 0$ . There is a unique Nash equilibrium in which trait 3 is inactive ( $q_3^* = 0$ ) and the equilibrium outcomes for the active traits are given by (59).*

**Proof** (1) The non-existence of Nash equilibria with two active traits follows directly from the three invasion conditions, and equations (63-64) characterize the unique solution to the system of equations  $(\Delta \mathbf{q}^*)_i = \pi^*$  for  $i \in \{1, 2, 3\}$  and  $\sum_i q_i^* = 1$ .

(2) Given  $\Psi_3 < 0$ , the point (63) lies outside the 3-dimensional simplex, and thus there is no Nash equilibrium with 3 active traits. Given that the other two invasion conditions are satisfied, there are no other Nash equilibria with two active traits.  $\square$

Thus, in the case where all traits are active, the  $\Psi$ 's also reflect the fraction of the population adopting each trait in equilibrium.

Given the existence of a unique Nash equilibrium in each case, we now establish the local stability of this equilibrium under the replicator dynamics. Because the population is characterized by the probability vector

$$\mathbf{q} = \begin{bmatrix} q_1 \\ q_2 \\ 1 - q_1 - q_2 \end{bmatrix}, \quad (65)$$

equation (17) becomes a two-dimensional dynamical system  $\{\dot{q}_1(q_1, q_2), \dot{q}_2(q_1, q_2)\}$ , which may be written in matrix form as

$$\begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \end{bmatrix} = \begin{bmatrix} q_1 & 0 \\ 0 & q_2 \end{bmatrix} \left[ \begin{bmatrix} 0 & \Delta_{12} & \Delta_{13} \\ \Delta_{21} & 0 & \Delta_{23} \end{bmatrix} \mathbf{q} - \mathbf{q}' \Delta \mathbf{q} \right]. \quad (66)$$

To assess local stability of a rest point  $\mathbf{q}^*$  using the standard approach, we first compute the Jacobian matrix

$$\mathbf{J}(\mathbf{q}) = \begin{bmatrix} \partial \dot{q}_1 / \partial q_1 & \partial \dot{q}_1 / \partial q_2 \\ \partial \dot{q}_2 / \partial q_1 & \partial \dot{q}_2 / \partial q_2 \end{bmatrix}_{\mathbf{q}} \quad (67)$$

and then evaluate this matrix at  $\mathbf{q}^*$  to obtain  $\mathbf{J}(\mathbf{q}^*)$ . If both eigenvalues of  $\mathbf{J}(\mathbf{q}^*)$  are negative, then  $\mathbf{q}^*$  is locally stable. For any square matrix, the sum of its eigenvalues equals its trace, and the product of its eigenvalues equals its determinant. To establish local stability of the Nash equilibrium in case (1), where  $\mathbf{q}^*$  is given by (63), it thus suffices to show that

$$\text{trace}(\mathbf{J}(\mathbf{q}^*)) = -\frac{\Delta_{12}\Delta_{23}\Delta_{31} + \Delta_{13}\Delta_{32}\Delta_{21}}{\Psi_1 + \Psi_2 + \Psi_3} < 0 \quad (68)$$

and

$$\det(\mathbf{J}(\mathbf{q}^*)) = \frac{\Psi_1\Psi_2\Psi_3}{\Psi_1 + \Psi_2 + \Psi_3} > 0. \quad (69)$$

For case (2), where  $\mathbf{q}^*$  is given by (59), the two eigenvalues of  $\mathbf{J}(\mathbf{q}^*)$  are

$$\lambda_1 = -\frac{\Delta_{12}\Delta_{21}}{\Delta_{12} + \Delta_{21}} \quad \text{and} \quad \lambda_2 = \frac{\Psi_3}{\Delta_{12} + \Delta_{21}}. \quad (70)$$

Thus, local stability of the Nash equilibrium is strict when  $\Psi_3 < 0$ , while this equilibrium is borderline stable when  $\Psi_3 = 0$ .

For each case, it is possible to establish global stability of the Nash equilibrium graphically using phase diagrams. But this is unnecessary because global stability follows immediately from more general results on replicator systems.<sup>16</sup> Consider

---

<sup>16</sup>Zeeman (1980) and Bomze (1983) provide exhaustive classifications of phase diagrams for replicator systems when  $n = 3$ . These analyses show that, for every robust (structurally stable) class with a unique locally stable rest point, this rest point is also globally stable. Cases (1) and (2) correspond to the phase diagrams labeled 7 and 35 in Bomze (1983), and to the phase diagrams labeled  $10_1$  and  $-10_2$  (with flows reversed) in Zeeman (1980).

case (1). We have just shown that the Nash equilibrium (which lies in the interior of the 3-dimensional simplex) is locally stable. It is straightforward (using the same approach) to show that the other 6 rest points of the replicator dynamics (which lie on the boundary of the simplex) are locally unstable. Further, given  $n = 3$ , the replicator dynamics admit no isolated periodic orbits (Hofbauer and Sigmund 1998, Theorem 4.2.1).<sup>17</sup> Non-isolated periodic orbits (like those generated by the rock-paper-scissors payoff matrix<sup>18</sup>) cannot arise when local stability of the interior rest point is strict. Given a unique locally stable rest point which lies in the interior of the simplex, and having eliminated the possibility of limit cycles, this rest point must be globally stable.<sup>19</sup> Now consider case (2). The Nash equilibrium (which lies on one edge of the simplex) is locally stable, and it is easily verified that other 5 rest points of the replicator dynamics (which also lie on the boundary of the simplex) are locally unstable. Given there is no interior rest point, all trajectories must converge to the boundary of the simplex (Hutson and Moran 1982), and the Nash equilibrium must be globally stable.

## 6 Permanence

Increasing the number of traits beyond  $n = 3$ , the dynamics and equilibria of the Bisin-Verdier model become more complicated. For instance, it is possible to have multiple stable rest points when  $n = 4$  (recall the example in 48) and stable limit cycles when  $n = 5$ .<sup>20</sup> Thus, when  $n$  is large, it appears impractical to attempt a full characterization of the stable attractors. However, for substantive purposes, perhaps the interesting question is whether any traits will become “extinct.” That is, given an initial condition in the interior of the  $n$ -dimensional simplex (so that  $q_i(0) > 0$  for all  $i$ ), will any trait ever “die out” in the long run (so that  $q_i^* = 0$  for at least one

<sup>17</sup>This result is proven for the generalized Lotka-Volterra model, but see Hofbauer and Sigmund (1998, Chap 7.5) on the equivalence between the Lotka-Volterra and replicator models.

<sup>18</sup>See, e.g., Hofbauer and Sigmund (1998, Chap 7) or Sandholm (2008, Chap 4).

<sup>19</sup>One might conjecture that, for case (1), the population game is stable, thus establishing global stability in this manner. However, this conjecture is false. For instance, given the payoff matrix

$$\Delta = \begin{bmatrix} 0 & \epsilon & 1 \\ \epsilon & 0 & \epsilon \\ \epsilon & \epsilon & 0 \end{bmatrix}$$

where  $\epsilon$  is small (e.g.,  $\epsilon = 1/8$ ), all three invasion conditions hold, and there is a unique, globally stable Nash equilibrium with  $q_1^* = 1/(1 + 2\epsilon)$  and  $q_2^* = q_3^* = \epsilon/(1 + 2\epsilon)$ . However, the stability condition (35) is not satisfied.

<sup>20</sup>For example, consider the *hypercycle system* where the payoff matrix  $\Delta$  is specified so that  $\Delta_{i,i+1} = 1$  for  $i \in \{1, \dots, n-1\}$ ,  $\Delta_{n1} = 1$ , and  $\Delta_{ij} = 0$  otherwise (Hofbauer and Sigmund 1998, Chap 12; Sandholm 2008, Chap 8). This payoff matrix generates a unique interior rest point ( $q_i^* = 1/n$  for all  $i$ ) which is locally stable for  $n \leq 4$  but locally unstable for  $n \geq 5$ . In the latter case, there is a stable limit cycle around the rest point.

*i*)? Recognizing that this question has already received considerable attention from mathematical biologists (see Hofbauer and Sigmund 1998, Chaps 12-14), the present section reviews and illustrates some key results.

Following Schuster et al (1979), the replicator system (17) is *permanent* when the boundary of the  $n$ -dimensional simplex is repelling.<sup>21</sup> A sufficient condition for permanence is given by the following

**Theorem 3 (Jansen 1987)** *The replicator system is permanent if there exists some  $\mathbf{p}$  in the interior of the  $n$ -dimensional simplex such that*

$$\mathbf{p}'\Delta\mathbf{x} > \mathbf{x}'\Delta\mathbf{x} \quad (71)$$

for all rest points  $\mathbf{x}$  on the boundary of the simplex.

See also Hofbauer and Sigmund (1998, Theorem 13.6.1). Rewriting (71) as

$$\sum_{i:x_i=0} p_i[(\Delta\mathbf{x})_i - \mathbf{x}'\Delta\mathbf{x}] > 0, \quad (72)$$

the scalar  $[(\Delta\mathbf{x})_i - \mathbf{x}'\Delta\mathbf{x}]$  may be interpreted as a “transversal” eigenvalue which is positive when the replicator system “calls for” the missing trait  $i$  (Hofbauer and Sigmund 1998, Chap 13.4).

To illustrate the application of this theorem, consider again the case where  $n = 3$ . Because there are three rest points on the edges of the simplex and three more at the corners, (71) is a system of 6 inequalities. For the rest point in (59), where

$$\mathbf{x}' = \left[ \frac{\Delta_{12}}{\Delta_{12} + \Delta_{21}} \quad \frac{\Delta_{21}}{\Delta_{12} + \Delta_{21}} \quad 0 \right], \quad (73)$$

we obtain the inequality

$$p_3[(\Delta\mathbf{x})_3 - \mathbf{x}'\Delta\mathbf{x}] = p_3 \left( \frac{\Psi_3}{\Delta_{12} + \Delta_{21}} \right) > 0 \quad (74)$$

which (for any choice of  $p_3$ ) reduces to the invasion condition  $\Psi_3 > 0$ .<sup>22</sup> In a similar way, given each of the other two rest points on the edges of the simplex, we obtain the invasion conditions  $\Psi_1 > 0$  and  $\Psi_2 > 0$ . Turning to the rest points at the corners of the simplex,  $\mathbf{x} = [1 \ 0 \ 0]'$  is associated with the inequality

$$p_2[(\Delta\mathbf{x})_2 - \mathbf{x}'\Delta\mathbf{x}] + p_3[(\Delta\mathbf{x})_3 - \mathbf{x}'\Delta\mathbf{x}] = p_2\Delta_{21} + p_3\Delta_{31} > 0. \quad (75)$$

This inequality, along with the analogous inequalities for the other two corners, is obviously satisfied for any choice of  $\mathbf{p}$ . Thus, for the  $n = 3$  case, permanence is

---

<sup>21</sup>Somewhat more formally, the system is permanent when there exists a compact set  $K$  in the interior of the simplex such that all trajectories in the interior end up in  $K$ .

<sup>22</sup>Note that the transversal eigenvalue in (74) corresponds to the eigenvalue labeled  $\lambda_2$  in (70).

guaranteed by Jansen’s Theorem when the three invasion conditions hold.<sup>23</sup> Moving beyond the 3-trait case, the number of “Jansen’s inequalities” expands rapidly in  $n$ . For instance, for  $n = 4$ , there are inequalities corresponding to each of the 4 corners of the 3-dimensional simplex, each of the 6 edges, and up to 4 faces (depending whether there is a rest point in the interior of the face).

Mathematical biologists have also established an important necessary condition for permanence, given by the following

**Theorem 4 (Hutson and Moran 1982)** *The replicator system is permanent only if the interior of the  $n$ -dimensional simplex contains a unique rest point  $\mathbf{q}^*$ .*

See also Hofbauer and Sigmund (1998, Theorem 13.5.1). Recognizing that any interior rest point of the replicator dynamics must be a Nash equilibrium, the rest point  $\mathbf{q}^*$  must be the unique solution to the system of  $(n + 1)$  equations

$$(\Delta \mathbf{q}^*)_i = \pi^* \quad \forall i \quad \text{and} \quad \sum_i q_i^* = 1. \quad (76)$$

Thus, if the solution to this system of equations lies outside the  $n$ -dimensional simplex, the replicator system is not permanent. Given  $n = 3$ , the solution to (76) is given by (63), and hence the three invasion conditions are also necessary for permanence.

While Theorems 3 and 4 hold for any payoff matrix  $\Delta$ , mathematical biologists have also considered permanence under the precise restrictions imposed by the Bisin-Verdier model ( $\Delta_{ii} = 0 \quad \forall i$ , and  $\Delta_{ij} \geq 0 \quad \forall j \neq i$ ). Biologists refer to the replicator system (17) under this restriction as a “catalytic network” and have studied these networks in the context of prebiotic evolution (Eigen and Schuster 1979). On this view, “traits” are types of macromolecules,  $\Delta_{ij}$  is positive when molecules of type  $j$  serve as catalysts for molecules of type  $i$ , and permanence is a property of self-sustaining (i.e., autocatalytic) networks.<sup>24</sup>

Because catalytic networks are sparse (in the sense that  $\Delta_{ij} = 0$  for most pairs  $\{i, j\}$ ), it is natural to adopt a graph-theoretic perspective. That is, we may construct a directed graph  $\Gamma(\Delta)$  in which the nodes are types of molecules (traits) and there is a directed edge from  $j$  to  $i$  when  $\Delta_{ij} > 0$ . Much of the theoretical analysis of catalytic networks has thus addressed whether permanence can be determined by properties of this graph. For instance,

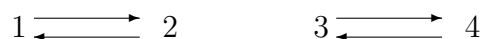
**Theorem 5 (Sigmund and Schuster 1984)** *If the catalytic network is permanent, then its graph  $\Gamma(\Delta)$  is strongly connected. That is, the matrix  $\Delta$  is irreducible.*

---

<sup>23</sup>Of course, this was already apparent from our finding that the (interior) Nash equilibrium in case (1) of Proposition 3 is globally stable.

<sup>24</sup>The simplest instance of a catalytic network is the hypercycle system discussed in footnote 20.

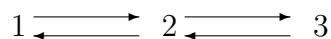
See also Stadler and Schuster (1996, Theorem 6). To illustrate, consider the replicator system with payoff matrix (48) where  $\epsilon = 0$ . This system satisfies the necessary condition for permanence in Theorem 4 (recall the interior Nash equilibrium given in footnote 11), but can be shown to fail the sufficiency condition (Jansen’s inequalities) given in Theorem 3. Nevertheless, given Theorem 5, we can immediately see that this system is not permanent because its graph (shown below) has more than one component.



While irreducibility is necessary for permanence, mathematical biologists have established an even stronger condition when  $n$  is small.

**Theorem 6 (Amann 1989)** *If  $n \leq 5$ , then the graph of a permanent catalytic network is Hamiltonian. That is, the graph contains a closed circuit that visits every node exactly once.*

See also Hofbauer and Sigmund (1998, Theorem 14.4.1). To illustrate, consider the catalytic network depicted below.



Although this network is strongly connected, it does not contain a Hamiltonian circuit, and hence it is not permanent.<sup>25</sup> Given  $n = 3$ , the necessity of a Hamiltonian circuit is evident from the population average payoff (64) when the system is permanent (case 1 of Proposition 3). The two terms in this expression,  $\Delta_{12}\Delta_{23}\Delta_{31}$  and  $\Delta_{13}\Delta_{32}\Delta_{21}$ , correspond to the two Hamiltonian circuits possible on the set of three nodes. If there is no Hamiltonian circuit, then both of these terms are zero, and hence the population average payoff is zero. But this contradicts permanence.<sup>26</sup>

Given our focus on intergenerational cultural transmission, the applicability of these last two theorems on catalytic networks may be limited because the payoff matrix  $\Delta$  is unlikely to be sparse. Indeed, given our maintained assumption in Sections 4 and 5 that all cultural distastes are strictly positive, the graph  $\Gamma(\Delta)$  is

---

<sup>25</sup>More generally, we may define a *plait* as a catalytic network with  $\Delta_{i,i+1} > 0$  and  $\Delta_{i+1,i} > 0$  for all  $i \in \{1, \dots, n-1\}$ , and  $\Delta_{ij} = 0$  otherwise. Stadler and Schuster (1996, Theorem 9) show that a plait is never permanent for  $n > 2$ .

<sup>26</sup>For  $n = 6$ , it is possible to construct examples of catalytic networks that are permanent but not Hamiltonian. See, e.g., Hofbauer and Sigmund (1998, Exercise 14.4.2).

complete. Moreover, as concluded by Stadler and Schuster (1996, p 131), there is only so much we can reasonably expect to learn about the dynamics of a replicator system from this (binary) graph-theoretic representation. Nevertheless, the analogy to catalytic networks offers some useful insight into the permanence of cultural traits. Consider traits  $i$  and  $j$  where  $\Delta_{ij}$  is large, so that  $i$  has a strong cultural distaste for  $j$ . Given the opportunity, parents with trait  $i$  might prefer to eliminate trait  $j$  from the population, thereby removing this pool of potential cultural parents. But adopting the biological perspective, trait  $j$  is a “catalyst” that may well be crucial for the long-run survival of trait  $i$ .<sup>27</sup> Of course, to maintain equilibrium, trait  $j$  must itself be catalyzed either directly by trait  $i$  or indirectly through some longer chain involving  $i$ , since the set of traits is permanent (“autocatalytic”) only if every trait indirectly catalyzes every other trait.

## 7 Concluding remarks

Generalizing the Bisin and Verdier (2000, 2001) model to permit more than two cultural traits, we have shown that the equilibrium outcome can change qualitatively as the number of traits rises. Given only two traits, the population converges to a unique limiting distribution, with a positive fraction of the population holding each trait. But given three traits, it is possible that one of the traits is “too tolerant” to remain viable. Thus, while minorities do “try harder” to socialize their children as their population share falls (as emphasized by Bisin and Verdier 2000, 2001), they may not try “hard enough” for their trait to survive if its cultural distaste for the other traits is weak. Increasing further the number of traits, the model can generate multiple equilibria when  $n = 4$  and limit cycles when  $n = 5$ . Thus, the  $n$ -trait model permits a much wider range of long-run outcomes than might have been anticipated from prior analyses of the 2-trait case.

But beyond providing these insights into the  $n$ -trait case, the present paper also makes a more foundational contribution to the economic analysis of intergenerational cultural transmission. Recasting the Bisin-Verdier model as an evolutionary game, our analysis illustrates a novel approach to studying intergenerational transmission which proceeds as if individuals choose traits directly, and the payoffs to traits correspond to cultural distastes. Perhaps more importantly, this new perspective allows research on intergenerational transmission to exploit the extensive literature on evolutionary game theory. In part, that literature provides elegant methods for addressing existing concerns, such as the global stability of long-run outcomes. But that literature may also suggest new questions, as illustrated by our focus on the long-run survival (permanence) of a set of traits. Given the analogy between catalytic

---

<sup>27</sup>More concretely, consider the title of Christian Smith’s (1998) book, *American Evangelicalism: Embattled and Thriving*. The current model suggests that evangelicals are thriving precisely because they feel besieged (and hence are “catalyzed”) by mainstream culture.

networks and systems of cultural traits, further engagement with the mathematical biology literature may prove especially fruitful.

The present analysis suggests a variety of questions for future research. For instance, given the apparent evolutionary benefits of strong cultural distastes, what limits the strength of these distastes? Casual empiricism indicates that, for many societies, the most “intolerant” traits are not always the most prominent. Reflection on this question might suggest the need to modify the simple two-stage socialization process assumed in the Bisin-Verdier model. If oblique socialization involves drawing *multiple* cultural parents from the population, and children adopt the trait of the cultural parent who is least hostile toward their biological parent, this might limit the advantages of strong distastes. Of course, many other generalizations of the two-stage process are possible (see Bisin and Verdier 2001 for other examples), and formal specification and analysis of these alternatives presents an interesting task for future research. Such extensions may well sever the tight link between the Bisin-Verdier model and the replicator dynamics. But as long as our current specification of the socialization process remains a special case, perhaps the literature on replicator systems will continue to provide some guidance.<sup>28</sup>

---

<sup>28</sup>Our key observation in this paper – that the Bisin-Verdier model is equivalent to a replicator system – is important primarily because much is already known about replicator systems. Even if future extensions break this link, researchers may still be able to draw upon existing results and the theoretical methods used to prove those results.

## References

- Amann, E (1989) *Permanence of Catalytic Networks*. Dissertation, University of Vienna.
- Bisin, Alberto, Giorgio Topa, and Thierry Verdier (2004) “Religious Inter-marriage and Socialization in the United States,” *Journal of Political Economy* 112:615-644.
- (2008) “Cultural Transmission, Socialization and the Population Dynamics of Multiple State Traits Distributions,” *International Journal of Economic Theory*, forthcoming.
- Bisin, Alberto and Thierry Verdier (2000) “Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits,” *Quarterly Journal of Economics* 115:955-988.
- (2001) “The Economics of Cultural Transmission and the Dynamics of Preferences,” *Journal of Economic Theory* 97:298-319.
- Bomze, Immanuel M (1983) “Lotka-Volterra Equations and Replicator Dynamics: A Two-Dimensional Classification,” *Biological Cybernetics* 48:201-211.
- Cavalli-Sforza, Luigi, and Marcus Feldman (1981) *Cultural Transmission and Evolution: A Quantitative Approach* Princeton: Princeton University Press.
- Eigen, M and P Schuster (1979) *The Hypercycle*. Berlin-Heidelberg: Springer.
- Hofbauer, Josef and William H Sandholm (2008) “Stable games and their dynamics,” unpublished manuscript, University of Vienna and University of Wisconsin.
- Hofbauer, Josef and Karl Sigmund (1998) *Evolutionary Games and Population Dynamics*. New York: Cambridge University Press.
- Hout, Michael, and Joshua R Goldstein (1994) “How 4.5 Million Irish Immigrants Become 40 Million Irish Americans,” *American Sociological Review* 59:64-82.
- Hout, Michael, Andrew Greeley, and Melissa J Wilde (2001) “The Demographic Imperative in Religious Change in the United States,” *American Journal of Sociology* 107:468-500.
- Hutson, V and W Moran (1982) “Persistence of Species Obeying Difference Equations,” *Bulletin of Mathematical Biology* 15:203-212.
- Jansen, Wolfgang (1987) “A Permanence Theorem for Replicator and Lotka-Volterra Systems,” *Journal of Mathematical Biology* 25:411-422.

- Samuelson, Larry (2002) "Evolution and Game Theory," *Journal of Economic Perspectives* 16:47-66.
- Sandholm, William H (2001) "Potential Games with Continuous Player Sets," *Journal of Economic Theory* 97:81-108
- (2008) *Population Games and Evolutionary Dynamics*. Unpublished book manuscript, forthcoming MIT Press. Draft posted at [www.ssc.wisc.edu/~whs](http://www.ssc.wisc.edu/~whs)
- Schuster, P, K Sigmund, and R Wolff (1979) "Dynamical Systems under Constant Organization. Part 3: Cooperative and Competitive Behavior of Hypercycles," *Journal of Differential Equations* 32:357-368.
- Sigmund, K, and P Schuster (1984) "Permanence and Uninvadability for Deterministic Population Models," in *Stochastic Phenomena and Chaotic behavior in Complex Systems* (Synergetics, Vol. 21) edited by P Schuster. New York: Springer-Verlag.
- Smith, Christian (1998) *American Evangelicalism: Embattled and Thriving*. Chicago: University of Chicago Press.
- Stadler, Peter F and Peter Schuster (1996) "Permanence of Sparse Catalytic Networks," *Mathematical Biosciences* 131:111-133.
- Taylor, P D and L B Jonker (1978) "Evolutionary Stable Strategies and Game Dynamics," *Mathematical Biosciences* 40:145-156.
- Zeeman, E C (1980) "Population Dynamics from Game Theory," in *Global Theory of Dynamical Systems* (Lecture Notes in Mathematics 819). New York: Springer.