Equilibrium allele distribution in trading populations

Gilles Saint-Paul

March 9, 2006

1 Introduction

In Saint-Paul (2002), I consider the evolution of the gene pool in a population under alternative economic institutions, and shown that alleles that cannot survive natural selection under autarky can survive under trade, because individuals can specialize in activities so as to avoid the fitness disadvantages associated with these alleles. The results are based on a very simplified representation of sexual reproduction, with only one chromosome (instead of pairs of chromosomes), and only two loci that determine the individual's productivity at two activities that affect fitness.

This paper generalizes these results for a more general system of sexual reproduction, with an arbitrary number of chromosomes and loci. Its contribution is twofold. First, it provides a set of assumptions under which one can meaningfully state that some alleles dominate their alternatives and eventually eliminate them in the long-run. Second, it extends the results in Saint-Paul (2002), by characterizing the distribution of alleles for a trading population in a *long-run equilibrium (*LRE), defined as a a stationary distribution of alleles which is also an equilibrium in an economic sense.

The central result is that fitness-reducing alleles can survive in a trading population, provided their frequency is not too large. However, the greater the number of loci that matter for fitness, the more stringent the conditions under

which these alleles can survive. That means that in the long-run, we expect low alleles to survive only at a relatively small number of loci. Knowing more about the long-run distribution of alleles when their initial distribution does not satisfy the conditions for an LRE would involve analyzing the dynamics, which I do not do here but is an interesting topic for further research.

2 Notations and genetic properties of stationary populations

A genotype consists of an $n-$tuple $g = (p_1, ..., p_n)$, where $i = 1, ..., n$ denotes a particular locus, and $p_i \in \{0, 1, ..., K\}$ is interpreted as the number of alleles of the "high type" at locus i (in the actual world where chromosomes come by pairs, one has $K = 2$). Therefore, there are $K - p_i$ alleles of the "low type" at locus i. The set of possible genotypes is denoted by S. We will also denote by $g[i]$ the ith element of g.

2.1 The survival function

The survival rate of an individual only depends on its genotype, and is denoted by $\varphi(g)$. Note that the $\varphi(g)$ function is not independent of culture. The opportunity to trade and specialize will dramatically change the $\varphi(g)$ mapping. It is useful to introduce the genetic improvement operators $T_z^{(i)}$, which, for any genotype g such that $g[i] < K - z$, maps it into another genotype $T_z^{(i)}g$, defined by $T_z^{(i)}g[j] = g[j], j \neq i$, and $T_z^{(i)}g[i] = g[i] + z$. Note that $T_{z+1}^{(i)} = T_z^{(i)} \circ T_1^{(i)}$.

The survival function is *monotonic* at locus i if it satisfies

$$\varphi(T_1^{(i)}g) - \varphi(g) \geq 0, \forall g \tag{1}$$

Thus, having more of a high allele at locus i cannot increase mortality, everything else equal. Note that this assumes that the role played by an allele in mortality has the same sign regardless of what other alleles are present.

We will say that a locus i is *selective* if

$$\varphi(T_1^{(i)}g) - \varphi(g) > 0, \forall g. \tag{2}$$

2.2 The distribution of offsprings

We assume a quite general process for transmitting genes to offsprings, which in particular is compatible with real-world genetics. When genotypes g' and g'' mate, the fraction of their offsprings with genotype g is given by a probability distribution function $F_{g'g''}(g)$. We shall assume that it satisfies the following properties:

1. *Gene conservation*

$$\begin{aligned} \forall g \forall i, u_i(g', g'') &= \sum_{g \in S} F_{g'g''}(g) g[i] \\ &= \frac{1}{2}(g'[i] + g''[i]) \end{aligned} \tag{3}$$

This says that on average, the number of high alleles at locus i among offsprings, denoted by $u_i(g', g'')$, is equal to the its average between the two parents. For a given pair of parents, the average among actual offsprings will be different from the parental average. However, with a continuum of individuals, the law of large number will apply, and $u_i(g', g'')$ will be equal to the population average of the number of H-alleles at i among all offsprings of all couples with genotypes g' and g''.

2. *Allele independence*

$$\frac{F_{g'g''}(p_1, ..., \bar{p}, ...p_n)}{\sum_{g, g[i]=\bar{p}} F_{g'g''}(g)} = \frac{F_{g'g''}(p_1, ..., \tilde{p}, ...p_n)}{\sum_{g, g[i]=\tilde{p}} F_{g'g''}(g)} \tag{4}$$

This assumption tells us that, among offsprings with the same parental genotypes, the distribution of other genes among those who have the same number of high alleles at locus i, does not depend on that particular number. If that

property did not hold, having many good alleles at one locus could in principle be systematically correlated with having many bad alleles at another locus, and this complementarity could sustain a positive amount of mortality-increasing alleles in the long-run, or, conversely, eliminate mortality-reducing ones.

3. *Mixing*

For any g', g'', i, *for all* z *such that*

$$\max(K/2, g'[i]) + \max(K/2, g''[i]) - K \leq z \leq \min(K/2, g'[i]) + \min(K/2, g''[i]) \tag{5}$$

,*there exists* g *such that* $g[i] = z$, $\min(g'[j], g''[j]) \leq g[j] \leq \max(g'[j], g''[j])$ for $j \neq i$, *and* $F_{g'g''}(g) > 0$.

The RHS of (5) is the maximum number of H-alleles at locus i if one inherits $K/2$ alleles from each parent; the LHS is the minimum number of H-alleles. That assumptions says that for any number between these two bounds, there is a positive probability for a couple g', g'' to have an offspring with exactly that number. Furthermore, we can pick up that offspring such that at all other loci, its number of high alleles is between that of its two parents. Loosely speaking, that means that the distribution of offsprings spans all possible cases.

4. *Symmetry*

$$\forall g', g'', F_{g'g''}(.) \equiv F_{g''g'}(.) \tag{6}$$

5. *Monotonicity*

For any i,*any* g'', *any* g' *such that* $g'[i] < K$, *and any* g *such that* $g[i] = 0$

$$\sum_{z=0}^{u} F_{T_1^{(i)}g',g''}(T_z^{(i)}g) \leq \sum_{z=0}^{u} F_{g',g''}(T_z^{(i)}g); \forall u < K \tag{7}$$

$$\sum_{z=0}^{K} F_{T_1^{(i)}g',g''}(T_z^{(i)}g) = \sum_{z=0}^{K} F_{g',g''}(T_z^{(i)}g) \tag{8}$$

This assumption says that if instead of g', a genetically improved genotype at locus i mates with g'', then holding the alleles at other loci constant, the proportion of H-alleles at locus i improves in a first-order stochastic dominance sense: offsprings are more likely to have a higher umber of H-alleles at i. Formally, applying the $T_z^{(i)}$ operators starting from an initial genotype g such that $g[i] = 0$, allows to compute the marginal distribution of $g[i]$ among offsprings holding other alleles constant. The last equality says that the partial distribution of the genotype at all other loci except i is invariant when one mates with a genetic improvement of g' at i instead of g'.

2.3 Demographics

These assumptions allow to write down the demographic evolution equations of each genotype. We denote by N_t total population at date t and by $x_g(t)$ the fraction of people with genotype g. People mate randomly. There are $x_{g'}(t)x_{g''}(t)N_t$ matches of types g' and g'' at date t. They produce ν offpsrings, and a fraction $\varphi(g)$ of offsprings with genotype g reach maturity. Consequently, $x_g(t)$ evolves according to

$$x_g(t+1)N_{t+1} = N_t \nu \varphi(g) \sum_{g'} \sum_{g''} x_{g'}(t) x_{g''}(t) F_{g'g''}(g).$$

Adding all these equations across all possible genotypes we get that

$$\frac{N_{t+1}}{N_t} = \nu \left(\sum_g \varphi(g) \sum_{g'} \sum_{g''} x_{g'}(t) x_{g''}(t) F_{g'g''}(g) \right).$$

It is also useful to define the population frequency of high alleles at locus i :

$$h_i(t) = \sum_g x_g(t) g[i].$$

Note that if the gene conservation law holds, then one also has

$$h_i(t) = \sum_{g'} \sum_{g''} u_i(g', g'') x_{g'}(t) x_{g''}(t). \tag{9}$$

3 Elimination of less fit alleles

In this section, I provide the basic results regarding the elimination of less fit alleles. A first lemma, which derives from the random mating and mixing properties, states that if a genotype exists and if a high allele exists in the population at locus i, then we can find another genotype that differs from it only in that it is "improved" at locus i, unless, of course, the initial genotype has the maximum number of H-alleles at i.

LEMMA 1 – Assume the mixing property holds. Assume there exists a steady state, a locus i and a genotype g such that in that steady state, $x_g > 0$, $g[i] < K$, and $h_i > 0$. Then $x_{T_1^{(i)}g} > 0$.

PROOF – First note that because of random mating there exists a positive measure of matches between two arbitrary genotypes, provided these genotypes are in positive measure in the parent population.

If $g[i] > 0$, the mixing property applied at locus i implies that offsprings of g with itself include $T_1^{(i)}g$ with positive probability. Assume $g[i] = 0$. Since $h_i > 0$, there exists g' such that $g'[i] > 0$ and $x_{g'} > 0$. We can then iterate the mixing property, by looking at stage k at the mates between g and $g^{(k)}$, starting with $g^{(0)} = g'$. If at stage k, there exists $j \neq i$ such that $g^{(k)}[j] \neq g[j]$, say $g^{(k)}[j] > g[j]$, by applying the mixing property at locus j we know that among the offsprings between g and $g^{(k)}$, there exists one $g^{(k+1)}$ such that $F_{gg^{(k)}}(g^{(k+1)}) > 0$ – implying $x_{g^{(k+1)}} > 0$ in steady state – $g^{(k+1)}[j] = g^{(k)}[j] - 1$, and $\mid g^{(k+1)}[l] - g\,[l]\mid \leq \mid g^{(k)}[l] - g\,[l]\mid$. In other words, the "genetic distance" between $g^{(k)}$ and g strictly goes down with k. Once we have reached the stage where $g^{(k)}[j] = g[j]$ for all $j \neq i$, we apply the same procedure to locus i, until we have produced an offspring such that $g^{(k)}[j] = g[j], j \neq i$ and $g^{(k)}[i] = g[i] + 1$. At that stage $g^{(k)} = T_1^{(i)}g$. QED.

The following key result tells us that genes which increase mortality eventually disappear:

PROPOSITION 1 – Assume that one of these two conditions holds:

(i) locus i is selective, OR

(ii) φ is monotonic at i and there exists one genotype $\hat{g}$ such that $x_{\hat{g}} > 0$ in steady state, $\hat{g}[i] < K$, and $\varphi(T_1^{(i)}\hat{g}) > \varphi(\hat{g})$.

Assume (A3) and (A4) holds. Then in any steady state with $h_i > 0$, one must have $h_i = K$.

PROOF – The frequency of the high allele at i evolves according to

$$\begin{aligned} h_i(t+1) &= \sum_g x_g(t+1)g[i] \\ &= \frac{N_t}{N_{t+1}} \sum_g \upsilon\varphi(g) \sum_{g'} \sum_{g''} x_{g'}(t)x_{g''}(t)F_{g'g''}(g)g[i]. \end{aligned}$$

In steady state, we have that $N_{t+1}/N_t = n$,

$$x_g = \frac{\upsilon\varphi(g)}{n} \sum_{g'} \sum_{g''} x_{g'}x_{g''}F_{g'g''}(g), \tag{10}$$

and

$$h_i = \frac{\upsilon}{n} \sum_{g'} \sum_{g''} x_{g'}x_{g''} \sum_g \varphi(g)F_{g'g''}(g)g[i]. \tag{11}$$

The term $Q(g', g'') = \sum_g \varphi(g)F_{g'g''}(g)g[i]$ can be rewritten as follows:

$$Q(g', g'') = \sum_{z=0}^{K} z \sum_{g, g[i]=z} \varphi(g)F_{g'g''}(g).$$

That can be rewritten as:

$$Q(g', g'') = \sum_{g, g[i]=0} \sum_{z=0}^{K} z\varphi(T_z^{(i)}g)F_{g'g''}(T_z^{(i)}g).$$

This formula rests on the fact that all the genotypes such that $g[i] = z$ can be deducted by applying the transform $T_z^{(i)}$ to all genotypes such that $g[i] = 0$.

Furthermore, the allele independence property implies that for g such that $g[i] = 0$,

$$F_{g'g''}(T_z^{(i)}(g)) = F_{g'g''}(g)\frac{q_z(g', g'', i)}{q_0(g', g'', i)}, \tag{12}$$

where $q_z(g', g'', i) = \sum_{g, g[i]=z} F_{g'g''}(g)$ is the total fraction of genotypes with $g[i] = z$ among the offsprings of g' and g''.[1]Note that one must have

$$u_i(g', g'') = \sum_{z=0}^{K} z q_z(g', g'', i). \tag{13}$$

Hence:

$$Q(g', g'') = \sum_{g, g[i]=0} \frac{F_{g'g''}(g)}{q_0(g', g'', i)} \sum_{z=0}^{K} z\varphi(T_z^{(i)}g) q_z(g', g'', i).$$

Now, if locus i is selective, then $\varphi(T_z^{(i)}g)$ is *strictly increasing* in z. Consequently we have

$$\sum_{z=0}^{K} z\varphi(T_z^{(i)}g) q_z(g', g'') \geq \left(\sum_{z=0}^{k} z q_z(g', g'', i)\right)\left(\sum_{z=0}^{k} \varphi(T_z^{(i)}g) q_z(g', g'', i)\right). \tag{14}$$

This inequality rests on the fact that $\sum_{z=0}^{k} q_z(g', g'', i) = 1$. It holds with a strict inequality unless all the $q_z(g', g'', i)$ but one are equal to zero.

We now show that unless $h_i = 0$ or $h_i = K$, there exists a pair of genotypes (g', g'', i) such that $x_{g'} > 0$, $x_{g''} > 0$, and (14) strictly holds. First note that if $0 < h_i < K$, there exists a genotype g_0 such that $x_{g_0} > 0$, and $0 < g_0[i] < K$.[2] Next, note that if there exists g_0 such that $0 < g_0[i] < K$, the mixing property implies that for two parents of the same genotype g_0, there

[1] If $q_0(g', g'', i) = 0$, we can write down the same steps using the smallest value of z such that $q_z(g', g'', i) > 0$ as a benchmark.

[2] The only other possibility is to only have genotypes such that $g_1[i] = 0$ and such that $g_2[i] = K$, but random mating and mixing imply that they will produce offsprings such that $0 < g[i] < K$.

is a positive probability of having an offpsring g such that $g[i] = z$, for any z between $2\max(K/2, g_0[i]) - K$ and $2\min(K/2, g_0[i])$. As long as $K \geq 2$ and $0 < g_0[i] < K$, there are more than two values of z that satisfy that property. Consequently, there are at least two strictly positive values of $q_z(g_0, g_0, i)$, and one can take $g' = g'' = g_0$.

Thus, if $0 < h_i < K$, it must be that there exists a pair (g', g'') such that $x_{g'} > 0, x_{g''} > 0$, and (14) strictly holds.

Alternatively, consider the case where φ is monotonic. Then (14) also holds. Furthermore, assume there exists $\hat{g}$ such that $x_{\hat{g}} > 0$, $\hat{g}[i] < K$, and $\varphi(T_1^{(i)}\hat{g}) - \varphi(\hat{g}) > 0$. Let $\hat{z} = \hat{g}[i]$. Then (14) will hold with strict inequality for g', g'' such that $x_{g'} > 0, x_{g''} > 0$, $q_{\hat{z}}(g', g'', i) > 0$ and $q_{\hat{z}+1}(g', g'', i) > 0$. If $\hat{z} > 0$, taking $g' = g'' = \hat{g}$ and applying the mixing property to locus i, generates both offsprings with $g[i] = \hat{z}$ and $g[i] = \hat{z} + 1$, implying that $q_{\hat{z}}(\hat{g}, \hat{g}, i) > 0$ and $q_{\hat{z}+1}(\hat{g}, \hat{g}, i) > 0$. If $\hat{g}[i] = 0$, the Lemma implies that $x_{T_1^{(i)}\hat{g}} > 0$. Taking $g' = \hat{g}$ and $g'' = T_1^{(i)}\hat{g}$ then generates both offsprings with $g[i] = 0$ and $g[i] = 1$, implying that $q_0(\hat{g}, \hat{g}, i) > 0$ and $q_1(\hat{g}, \hat{g}, i) > 0$. Thus, we can again pick up a genotype g and a pair (g', g'') such that $x_{g'} > 0, x_{g''} > 0$, and (14) strictly holds.

From (14), we get that

$$\begin{aligned} Q(g', g'') &\geq \sum_{g, g[i]=0} \frac{F_{g'g''}(g)}{q_0(g', g'', i)} \left(\sum_{z=0}^{k} z q_z(g', g'', i) \right) \left(\sum_{z=0}^{k} \varphi(T_z^{(i)} g) q_z(g', g'', i) \right) \\ &= S(g', g'') \end{aligned} \tag{15}$$

Once again, there exists a pair (g', g'') such that $x_{g'} > 0, x_{g''} > 0$, and (15) strictly holds. The RHS can be rewritten

$$\begin{aligned} S(g', g'') &= u_i(g', g'') \sum_{g, g[i]=0} F_{g'g''}(g) \left(\sum_{z=0}^{k} \varphi(T_z^{(i)} g) \frac{q_z(g', g'', i)}{q_0(g', g'', i)} \right) \\ &= u_i(g', g'') \sum_{g} F_{g'g''}(g) \varphi(g), \end{aligned}$$

where the first step derives from (13) and the second one from (12).

Inequality (15) means that the fitness of the high alleles in the gene pool of the offsprings of g' and g'' is higher than the average fitness of the offsprings as individuals, because those with more H-alleles at i live longer. In order to get that, the allele invariance property is needed. Otherwise, it could be that the offsprings of g', g'' that have a high $g[i]$ have a lower fitness than the others because they are systematically poorly endowed at other loci.

Going back to (11), we see that

$$\begin{aligned} \frac{nh_i}{v} &= \sum_{g'}\sum_{g''} x_{g'}x_{g''}Q(g',g'') \\ &> \sum_{g'}\sum_{g''} x_{g'}x_{g''}S(g',g'') \\ &= \sum_{g'}\sum_{g''} x_{g'}x_{g''}u_i(g',g'')\sum_{g} F_{g'g''}(g)\varphi(g) \\ &= \tilde{h}_i \end{aligned} \tag{16}$$

where the strict inequality comes from the fact that $Q(g',g'') > S(g',g'')$ for at least one pair (g',g'') such that $x_{g'}x_{g''} > 0$.

We now have

$$\tilde{h}_i = \sum_{g',g'[i]=0}\sum_{z=0}^{k}\sum_{g''}\frac{1}{2}(z+g''[i])x_{T_z^{(i)}g'}x_{g''}R(g',g'',z), \tag{17}$$

where we have applied gene conservation and $R(g',g'',z)$ is defined as

$$R(g',g'',z) = \sum_{g} F_{T_z^{(i)}g',g''}(g)\varphi(g).$$

Observe that R can be rewritten as

$$R(g',g'',z) = \sum_{g,g[i]=0}\sum_{v=0}^{K} F_{T_z^{(i)}g',g''}(T_v^{(i)}g)\varphi(T_v^{(i)}g).$$

Furthermore, one can write $F_{T_z^{(i)}g',g''}(T_v^{(i)}g) = \Phi(v,z) - \Phi(v-1,z)$, where $\Phi(v,z) = \sum_{u=0}^{v} F_{T_z^{(i)}g',g''}(T_u^{(i)}g)$. Iterating the monotonicity property, we find that $\Phi(v,z)$ is nondecreasing in z, while $\Phi(K,z)$ does not depend on z. We then have that

$$
\begin{aligned}
& \sum_{v=0}^{K} F_{T_z^{(i)}g',g''}(T_v^{(i)}g)\varphi(T_v^{(i)}g) \\
= & \sum_{v=0}^{K} \left(S(v,z) - S(v-1,z)\right)\varphi(T_v^{(i)}g) \\
= & \sum_{v=0}^{K-1} S(v,z)\left[\varphi(T_v^{(i)}g) - \varphi(T_{v+1}^{(i)}g)\right] + S(K,z)\varphi(T_K^{(i)}g).
\end{aligned}
$$

Since φ is monotonic, the term in brackets is nonpositive. Thus, the sum is nondecreasing in z, while the last term is constant in z. Therefore, the LHS is nondecreasing in z, for any g such that $g[i] = 0$. Summing this property across these g's, we also find that $R(g', g'', z)$ is nondecreasing in z. Roughly, that property means that the average mortality of offsprings improves when one parent is genetically enhanced at locus i. The monotonicity property is needed to get that. Otherwise, it could be that parents with more H-alleles at i, everything else equal, have an $F_{g'g''}$ systematically biased toward high-mortality genotypes.

Let us now go back to (17), which we can rewrite

$$
\tilde{h}_i = \sum_{g''} x_{g''} \sum_{g',g'[i]=0} \sum_{z=0}^{k} \frac{1}{2}(z + g''[i]) x_{T_z^{(i)}g'} R(g', g'', z).
$$

For a given g'', we have that $\sum_{g',g'[i]=0} \sum_{z=0}^{k} x_{T_z^{(i)}g'} = 1$, that $\frac{1}{2}(z + g''[i])$ increases with z and that $R(g', g'', z)$ weakly increases with z. Thus, once again, we have the following inequality:

$$
\begin{aligned}
& \sum_{g',g'[i]=0} \sum_{z=0}^{k} \frac{1}{2}(z+g''[i]) x_{T_z^{(i)} g'} R(g',g'',z) \\
\geq & \left(\sum_{g',g'[i]=0} \sum_{z=0}^{k} \frac{1}{2}(z+g''[i]) x_{T_z^{(i)} g'} \right) \bullet \left(\sum_{g',g'[i]=0} \sum_{z=0}^{k} x_{T_z^{(i)} g'} R(g',g'',z) \right) \\
= & \frac{1}{2}(h_i + g''[i]) \left(\sum_{g',g'[i]=0} \sum_{z=0}^{k} x_{T_z^{(i)} g'} R(g',g'',z) \right).
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
\tilde{h}_i & \geq \sum_{g''} x_{g''} \frac{1}{2}(h_i + g''[i]) \left(\sum_{g',g'[i]=0} \sum_{z=0}^{k} x_{T_z^{(i)} g'} R(g',g'',z) \right) \\
& = \sum_{g''} x_{g''} \frac{1}{2}(h_i + g''[i]) \left(\sum_{g'} x_{g'} \sum_{g} F_{g'g''}(g) \varphi(g) \right) \\
& = \frac{1}{2} \frac{n h_i}{\upsilon} + \frac{1}{2} \sum_{g''} \sum_{g'} g''[i] x_{g''} x_{g'} \sum_{g} F_{g'g''}(g) \varphi(g) \frac{\mu(g)\nu}{\mu(g) + \nu - \bar{\mu}},
\end{aligned}
$$

where the steady-state condition (10)has been used to derive the first term.

By virtue of (16), (3) and (6), the last term in that formula must be equal to $\tilde{h}_i/2$, so that $\tilde{h}_i \geq \frac{n h_i}{\upsilon}$. (16) then implies that $h_i > h_i$, which is a contradiction. Hence, it must be that either $h_i = 0$ or $h_i = K$. Q.E.D.

The last set of inequalities tell us that since parents who have a greater $g'[i]$ have children with a higher fitness, these parents' children tend to increase the survival rate of the high allele at i relative to average. Since, in addition, the survival rate of the high allele at i among their children is greater than their children's average survival rate, these two effects together imply that the fitness of the high allele at i is strictly higher than average. But that cannot be in steady state, unless $\varphi_i = 0$ or K.

4 Autarky

We now describe how an individual's genotype g affects his/her productivity at various activities, depending on the ecomic setting.

The alleles present at a given locus i determines the individual's productivity at a corresponding activity denoted by the same index i. This productivity is a strictly increasing function $f_i(g[i])$ of $g[i]$, the number of H-alleles at locus i. Any individual has a total time endowment equal to 1. The time allocation constraint of genotype g is therefore given by

$$\sum_{i=1}^{n} \frac{v_i}{f_i(g[i])} \leq 1, \tag{18}$$

where v_i is the individual's output in activity i.

Finally the individual's fitness is

$$\varphi(g) = u(y_1, ..., y_n),$$

where y_i is the individual's consumption of activity i, and u is the "utility function", which is concave in each argument, and satisfies the "Inada conditions": $\lim_{y_i \longrightarrow 0} \frac{\partial u}{\partial y_i} = +\infty$, $\lim_{y_i \longrightarrow +\infty} \frac{\partial u}{\partial y_i} = 0$.

Under autarky, we have $y_i = v_i$, and the following result holds:

PROPOSITION 2 – Under autarky, all loci are selective. Therefore, in any steady state such that $h_i > 0, \forall i$, all individuals are of genotype $g_{\max}$, i.e. the H-allele is fixed at all locations.

Proof – Type $T_1^{(i)}g$ has a more favorable time budget constraint than type g. Therefore, it achieves a higher fitness. The rest follows from the previous subsection. Q.E.D.

Note that the case $h_i = 0$ is not of interest: it means that the high allele does not exist at that locus.

5 Trade

Let us now look at the trade case. Each good i is traded at price p_i. We assume the following normalization for the price vector (p_i)

$$\sum_{i=1}^{n} p_i = 1 \tag{19}$$

People allocate their time between the various activities so as to maximize their income $R(g) = \sum_{i=1}^{n} p_i v_i$, subject to the time allocation constraint (18). Their demand vector is the one which maximizes u subject to their budget constraint:

$$\sum_{i=1}^{n} p_i y_i = R(g)$$

Types with lower incomes must achieve lower fitness and therefore disappear in the long-run.

Furthermore, φ must be monotonic at all loci. The reason is that the vector $(v_1, ..., v_n)$ supplied by a genotype g can also be supplied by genotype $T_1^{(i)}g$. On the other hand, all loci need not be selective, as genotypes with fewer H-alleles at locus i may achieve the same income as fitter genotypes, by just specializing.

Define a long-run equilibrium (LRE), as a stationary state such that the economy is in equilibrium, i.e. each genotype sets its supply and demand as just described, and markets clear for each good. The following proposition generalizes the results derived for the two loci case in Saint-Paul (2002).

PROPOSITION 3 – (i) In any LRE such that $h_i > 0, \forall i$, a given type only supplies goods corresponding to loci in their genotype where they have the highest number of H–alleles: $v_i(g) > 0 \Longrightarrow g[i] = K$.

(ii) In any LRE such that $h_i > 0$, the price vector is $p^* = (p_1^*, ..., p_q^*)$ such that

$$p_i^* = \frac{1/f_i(K)}{\sum_j \frac{1}{f_j(K)}} \tag{20}$$

(iii) In any LRE, there exists a locus j such that $\varphi_j = K$, i.e. allele H is fixed at locus j.

Proof of (i) – Iterating the mixing property with appropriately chosen parents, one can easily show that if $h_i > 0, \forall i$, in steady state there exists a strictly positive supply of genotypes with a arbitrary, strictly positive number of H-alleles $g[i]$ at each locus i. In particular, there exists a strictly positive mass of the best genotype $g_{\max} : x_{g_{\max}} > 0$. Next, note that if $R(g') > R(g)$, then genotype g' achieves higher fitness than g, and hence $\varphi(g') > \varphi(g)$.

Assume there exists a genotype $\hat{g}$ such that $v_l(\hat{g}) > 0$ for l such that $\hat{g}[l] < K$. Clearly, the plan $(v_1(\hat{g}), ..., v_l(\hat{g})\frac{f_l(g[l]+1)}{f_l(g[l])}, ..., v_n(\hat{g}))$ achieves a strictly higher income level and is feasible (i.e. satisfies (18)) for $T_1^{(i)}g$. Consequently, $R(T_1^{(i)}g) > R(g)$, implying $\varphi(T_1^{(i)}g) > \varphi(g)$. But, given that φ is monotonic, Proposition 1, under assumption (ii), would then imply that $\varphi_i = K$, which makes it impossible for $\hat{g}$ to exist. Consequently, any type g only supplies goods where it has an H-allele.

Proof of (ii) – The price vector defined by (20) is the one which makes type $g_{\max}$ indifferent between all activities. Assume there exists an LRE with a different price vector. Then there exists a pair of goods (j, k) such that

$$\frac{p_j}{p_k} < \frac{f_k(K)}{f_j(K)}, \tag{21}$$

and $v_j(g_{\max}) = 0$ since more income is yielded for type $g_{\max}$ by offering good k than good j.

Since u satisfies the Inada conditions, the demand for good j is strictly positive; since $g_{\max}$ does not supply good j, there exists $g \neq g_{\max}$ such that $x_g > 0$ and $v_j(g) > 0$. By virtue of (i), $g[j] = K$. Furthermore, $g[k] < K$, otherwise g would prefer to supply k instead of j as well.

The income of type g is $R(g) = \sum_{i=1}^{q} p_{(i} v_i(g) = \sum_{i \neq j} p_i v_i(g) + p_j v_j(g)$. The supply vector $(v_i(g))$ is feasible for type $T^{(k)}_{K-g[k]}g$, since $T^{(k)}_{K-g[k]}g$ is more

productive than g at all activities. The supply vector (v'_i) defined by $v'_i = v_i(g)$, $i \neq j, k, v'_j = 0, v'_k = v_k(g) + v_j(g)\frac{f_k(K)}{f_j(K)}$ also satisfies (18) for $T^{(k)}_{K-g[k]}g$. Therefore,

$$\begin{aligned} R(T^{(k)}_{K-g[k]}g) &\geq \sum_{i=1}^{q} p_i v'_i \\ &= R(g) - p_j v_j(g) + p_k \frac{f_k(K)}{f_j(K)} v_j(g) \\ &> R(g), \end{aligned}$$

where the last inequality comes from (21). But, this cannot hold since it again implies $\varphi(T^{(k)}_{K-g[k]}g) > \varphi(g)$. Consequently, there exists $\hat{g} \in \{g, T^{(k)}_1 g, ...T^{(k)}_{K-g[k]}\}$ such that $\varphi(T^{(k)}_1 \hat{g}) > \varphi(\hat{g})$. Furthermore, as $x_g > 0$, iterating Lemma 1 implies that $x_{\hat{g}} > 0$. Monotonicity of φ then implies that (ii) in Proposition 1 is satisfied. Consequently, $h_k = K$. But that contradicts the requirement that $g[k] < K$. Q.E.D.

Proof of (iii) – Suppose not; then by iterating the mixing property with appropriately chosen parents, one can prove that $x_{g_{\min}} > 0$. But that contradicts (i).

Q.E.D.

The preceding proposition tells us what properties an LRE must necessarily have, but does not tell us whether an LRE exists and whether, as in the preceding analysis, one can construct equilibria with a positive level of some L-alleles. We now establish a result which tells us that an LRE exists with a strictly positive proportion of L-alleles, provided these alleles are not too frequent.

To do so, for any subset S of $\{1, ...n\}$ we define $\tilde{S}$ as $\tilde{S} = \{g, g[i] = K \Rightarrow i \in S\}$. $\tilde{S}$ is the set of all genotypes such that their loci saturated with H-alleles (which define the activities at which they can possibly specialize) are all in S.

PROPOSITION 4 – Let $Rw_i(p)$ be the inverse demand function for the fitness maximization problem of an individual with income R facing price vector

p. Let

$$D_i = \frac{1}{\sum_j \frac{1}{f_j(K)}} w_i(\mathbf{p^*})$$

Then there exists an LRE with a distribution $\{x_g\}$ *of genotypes if and only if this distribution satisfies the following property:*

$$\forall S \subset \{1, ...q\}, \sum_{i \in S} \frac{D_i}{f_i(K)} \geq \sum_{g \in \tilde{S}} x_g. \tag{22}$$

Proof – We first prove that this condition is necessary. The RHS of (22) is the total time supplied by genotypes in $\tilde{S}$ (relative to the population's total time). Proposition 3, (i) implies that it must be allocated among goods i such that $g[i] = K$, i.e. among goods in S. It also implies that in any candidate equilibrium, income per capita (equal to the income of any genotype) must be equal to $\frac{1}{\sum_j \frac{1}{f_j(K)}}$. Thus, D_i is the per capita amount of good i consumed and produced in any candidate equilibrium. The LHS of (22) is therefore the total time input needed to produce all the goods in S. It must be greater than or equal to its RHS, since genotypes in $\tilde{S}$ cannot produce any other good. Otherwise, supply would exceed demand. Note that (22) applied to $S = \emptyset$ implies that one H-allele is fixed (the LHS is then the total supply of all genotypes g such that $g[i] < K, \forall i$). Also, (22) applied to $S = \{1, ...n\}$ boils down to Walras' law, since it is equivalent to $\sum_{i=1}^{q} p_i^* w_i(p^*) \geq 1$, and by Walras law $\sum_{i=1}^{q} p_i^* w_i(p^*) = 1$.

Let us now prove sufficiency. In order to do so, we construct a set of functions $m_i(g)$, representing the share of time of genotype g devoted to activity i, such that:

$$m_i(g) > 0 \Longrightarrow g[i] = K \tag{23}$$

$$\sum_{g, g[i]=K} m_i(g) x_g = \frac{D_i}{f_i(K)}. \tag{24}$$

$$\sum_i m_i(g) = 1, \forall g \tag{25}$$

If we are able to construct such functions, then this is indeed an equilibrium, since supply equals demand for all goods, and since the price vector in (20) implies that a genotype is indifferent between supplying all the goods at which it has K H-alleles.

To construct the $m_i(g)$, we use the following algorithm. We start from any arbitrary allocation $m_i^{(0)}(g)$ satisfying (23) and (25). This defines the initial stage.[3] Then we move from stage (k) to stage $(k+1)$ as follows. At the beginning of stage (k), the set $\{1, ...n\}$ can be partitioned into three subsets:

$$\begin{aligned} H_0^{(k)} &= \{i, \sum_{g, g[i]=K} m_i^{(k)}(g) x_g = \frac{D_i}{f_i(K)}\} \\ H_+^{(k)} &= \{i, \sum_{g, g[i]=K} m_i^{(k)}(g) x_g < \frac{D_i}{f_i(K)}\} \\ H_-^{(k)} &= \{i, \sum_{g, g[i]=K} m_i^{(k)}(g) x_g > \frac{D_i}{f_i(K)}\}. \end{aligned}$$

That is, those goods for which supply equals demand, those for which there is excess demand, and those for which there is excess supply. Note that since $\sum_i D_i/f_i(K) = 1$, $H_+^{(k)}$ is empty if and only if $H_-^{(k)}$ is empty. If $H_+^{(k)} = H_-^{(k)} = \emptyset$, then we have an equilibrium, and the algorithm stops.

Assume therefore that it is not the case. Then neither $H_+^{(k)}$ nor $H_-^{(k)}$ are empty. We now distinguish two cases.

Case A. Assume there exists a partition (H_A, H_B) of $\{1, ...n\}$ such that

$$H_-^{(k)} \subseteq H_A; H_+^{(k)} \subseteq H_B.$$

and:

[3] One can trivially check that such an allocation exists, since one $H-$allele is fixed, all genotypes have at least one locus where $g[i] = K$.

$$\forall g, \left(\exists i \in H_A, m_i^{(k)}(g) > 0\right) \Longrightarrow (\forall i \in H_B,\ g[i] < K) \qquad \text{(P)}$$

That is, people who *do* produce goods in H_A *cannot* produce goods in H_B.

For any good i, let

$$G(i) = \{g, m_i^{(k)}(g) > 0\}.$$

Clearly, one has $g[i] = K$, for all $g \in G(i)$. We then have

$$\sum_{i \in H_A} \sum_{g \in G(i)} x_g m_i^{(k)}(g) > \sum_{i \in H_A} \frac{D_i}{f_i(K)}.$$

This strict inequality comes from the fact that $H_-^{(k)} \subseteq H_A \subseteq H_0^{(k)} \cup H_-^{(k)}$ and from the fact that $H_-^{(k)}$ is non-empty.

Furthermore,

$$\sum_{g \in \tilde{H}_A} x_g = \sum_{g \in \tilde{H}_A} x_g \sum_{i \in H_A} m_i^{(k)}(g).$$

This is because if $g \in \tilde{H}_A$ and $m_i^{(k)}(g) > 0$, then $g[i] = K$, implying $i \in \tilde{H}_A$. Therefore $\sum_{i=1}^n m_i^{(k)}(g) = \sum_{i \in H_A} m_i^{(k)}(g) = 1$.

Interverting, we get

$$\sum_{g \in \tilde{H}_A} x_g = \sum_{i \in H_A} \sum_{g \in \tilde{H}_A} x_g m_i^{(k)}(g)$$

Now, note that $G(i) \subseteq \tilde{H}_A$: If I produce one good in H_A, all my loci with K H-alleles are also in H_A. Consequently,

$$\sum_{g \in \tilde{H}_A} x_g \geq \sum_{i \in H_A} \sum_{g \in G(i)} x_g m_i^{(k)}(g) > \sum_{i \in H_A} \frac{D_i}{f_i(K)},$$

which clearly violates assumption (22). Case A is therefore ruled out.

Case B. Assume then that there exists no such partition. We can construct a chain of q goods $i_1, ..., i_q$ such that the following property holds:

PROPERTY Q:

(a) $i_1 \in H_-^{(k)}$;

(b) $i_q \in H_+^{(k)}$;

(c) $i_l \in H_0^{(k)}$, for $l = 2, ..., q-1$.

(d) $\forall l \in \{1, ...q-1\}, \exists g_l$, $m_{i_l}^{(k)}(g_l) > 0$ and $g_l[i_{l+1}] = K$.

To construct such a chain, proceed as follows. We will write $i \blacktriangleright i'$ if: ($\exists g$, $m_i^{(k)}(g) > 0$ and $g[i'] = K$). In this case, we call $\Omega(i, i')$ the set of genotypes that satisfy this property: $\Omega(i, i') = \{g,\ m_i^{(k)}(g) > 0$ and $g[i'] = K\}$. Property (d) implies that the chain we want to construct is such that $i \blacktriangleright i'$.

Start from a set $H_A^{(0)} = H_-^{(k)}$. As property (P) is violated, there exists $i_1' \in H_A^{(0)}, i_2' \in \{1, ..., n\} - H_A^{(0)}$, such that $i_1' \blacktriangleright i_2'$.

If $i_2' \in H_+^{(k)}$, stop the procedure there, and take $i_1 = i_1', i_2 = i_2', g_1 \in \Omega(i_1, i_2)$.

If not, then $i_2' \in H_0^{(k)}$. Add i_2' to H_A: $H_A^{(1)} = H_A^{(0)} \cup \{i_2'\}$. Since $H_0^{(k)} \cap H_A^{(1)} = \{i_2'\}$,it must be that $i \in H_A^{(1)} \cap H_0^{(k)} \Longrightarrow \exists j \in H_-^{(k)}$, $j \blacktriangleright i$. Use again the fact that (P) is violated. There exists $i_1'' \in H_A^{(1)}$, $i_2'' \in \{1, ..., n\} - H_A^{(1)}$, such that $i_1'' \blacktriangleright i_2''$. Given that either $i_1'' = i_2'$ or $i_1'' \in H_-^{(k)}$, there exists a chain of length q, $(i_1^{(1)}, ...i_q^{(1)} = i_2'')$ such that $i_1^{(1)} \in H_-^{(k)}$, $i_l^{(1)} \epsilon H_A^{(1)} \cap H_0^{(k)}, l < q$, and $i_l^{(1)} \blacktriangleright i_{l+1}^{(1)}$. If $i_2'' \in H_+^{(k)}$, we use that chain and stop the procedure. Otherwise, we add i_2'' to H_A, and iterate again.

More generally, at each iteration (r), there is a set $H_A^{(r)}$ such that $H_-^{(k)} \subseteq H_A^{(r)} \subseteq H_0^{(k)} \cup H_-^{(k)}$, and such that for all $i \in H_A^{(r)} \cap H_0^{(k)}$, there exists a chain $(i_1', ...i_q')$ such that $i_1' \in H_-^{(k)}, i_l' \in H_A^{(r)} \cap H_0^{(k)}, l > 1$, $i_q' = i$, and $i_l' \blacktriangleright i_{l+1}'$ (The chain property). Because (P) is violated, there exists $i_1'' \in H_A^{(r)}$ and $i_2'' \in \{1, ..., n\} - H_A^{(r)}$, such that $i_1'' \blacktriangleright i_2''$. Let $(i_1''', ...i_q''' = i_1'')$ be the chain corresponding to i_1''. If $i_2'' \in H_+^{(k)}$, we use the chain $(i_1''', ...i_q''' = i_1'', i_2'')$ and stop the procedure. If $i_2'' \in H_0^{(k)}$, we use $H_A^{(r+1)} = H_A^{(r)} \cup \{i_2''\}$ and iterate the procedure. As the new member i_2'' is connected to i_1''' via the chain $(i_1''', ...i_q''' = i_1'', i_2'')$, $H_A^{(r+1)}$ still satisfies the chain property. As $i_2'' \in H_0^{(k)}$, it also satisfies $H_-^{(k)} \subseteq H_A^{(r+1)} \subseteq H_0^{(k)} \cup H_-^{(k)}$. As the number of elements in $H_A^{(r+1)}$ goes up

by one unit at each iteration, one must find an i_2'' in $H_+^{(k)}$ a finite number of iterations.

Next, we can use such a chain to construct a new allocation of labor for stage $k+1$. Let

$$\Delta_0 = \min(m_{i_l}^{(k)}(g_l), l = 1, ..q-1\},$$

and

$$\Delta = \min(\sum_{g \in G(i_1)} m_{i_1}^{(k)}(g)x_g \ - \frac{D_{i_1}}{f_{i_1}(K)}, \frac{D_{i_q}}{f_{i_q}(K)} - \sum_{g \in G(i_q)} m_{i_q}^{(k)}(g)x_g, \Delta_0)$$

Define the new allocation as follows:

$$\begin{aligned} m_{i_l}^{(k+1)}(g_l) &= m_{i_l}^{(k)}(g_l) - \Delta; \\ m_{i_{l+1}}^{(k+1)}(g_l) &= m_{i_{l+1}}^{(k)}(g_l) + \Delta; \\ m_i^{(k+1)}(g_l) &= m_i^{(k)}(g_l), i \neq i_1, ...i_q \\ \forall_i, m_i^{(k+1)}(g) &= m_i^{(k)}(g), \ g \neq g_1, ..., g_{q-1}. \end{aligned}$$

The new allocation clearly still satisfies (23) (as $g_l[i_{l+1}] = K$), and (25) (as $m_{i_l}^{(k+1)}(g_l) + m_{i_{l+1}}^{(k+1)}(g_l) = m_{i_l}^{(k)}(g_l) + m_{i_{l+1}}^{(k)}(g_l)$). Futhermore, one has $m_i^{(k+1)}(g) \geq 0$, as $\Delta \leq \Delta_0$. Finally, for $1 < l < q$, $m_{i_l}^{(k+1)}(g_l) + m_{i_l}^{(k+1)}(g_{l-1}) = m_{i_l}^{(k)}(g_l) + m_{i_l}^{(k)}(g_{l-1})$. Hence, all markets that were in equilibrium remain so. Furthermore, as $\Delta \leq \sum_{g \in G(i_1)} m_{i_1}^{(k)}(g)x_g \ - \frac{D_{i_1}}{f_{i_1}(K)}$, market i_1 weakly remains in excess supply, and similarly market i_q weakly remains in excess demand. Therefore:

$$H_0^{(k)} \subseteq H_0^{(k+1)};$$

$$H_-^{(k)} \supseteq H_-^{(k+1)};$$

$$H_+^{(k)} \supseteq H_+^{(k+1)}.$$

Finally, we note that either

(i) $H_0^{(k)} \subsetneq H_0^{(k+1)}$, which will be true provided $\Delta = \sum_{g \in G(i_1)} m_{i_1}^{(k)}(g) x_g - \frac{D_{i_1}}{f_{i_1}(K)}$ or $\Delta = \frac{D_{i_q}}{f_{i_q}(K)} - \sum_{g \in G(i_q)} m_{i_q}^{(k)}(g) x_g$. In such cases the new allocation restores equilibrium in market i_1 (resp. i_q)

(ii) Or, the chain $(i_1, ...i_q)$ and its associated chain of genotypes $(g_1, ...g_{q-1})$ no longer satisfy Q; that is the case if $\Delta = m_{i_l}^{(k)}(g_l)$ for some l, in which case $m_{i_l}^{(k+1)}(g_l) = 0$. In such a case, we have constructed a new allocation such that $\sum_g \mid \{i, m_i(g) = 0\} \mid$ has increased by at least one unit, and which satisfies (23) and (25).

Thus, at each stage, the quantity $\sum_g \mid \{i, m_i(g) = 0\} \mid + \mid H_0^{(k)} \mid$ strictly increases. As it is bounded, the procedure cannot go on forever, and the only case in which one cannot iterate it is if $H_+^{(k)} = H_-^{(k)} = \emptyset$. This proves the existence of equilibrium. Q.E.D.

Clearly, conditions (22) are pretty stringent, so that it is not straightforward to construct an equilibrium. However for $x_{g_{\max}}$ close enough to 1, i.e. x_g small enough when $g \neq g_{\max}$, they are clearly satisfied, since $x_{g_{\max}}$ appears on the RHS only for $S = \{1, ...q\}$, in which case (22) is always satisfied with equality, due to Walras' law: $\sum_{i=1}^q \frac{D_i}{z_{Hi}} = \sum_{i=1}^q p_i^* f_i(\mathbf{p}^*) = 1 = \sum_g x_g$. Therefore there always exist equilibria with a strictly positive fraction of genotypes with L-alleles, provided this fraction is small enough.

Note that the greater the number of loci, the greater the number of conditions that must hold. Intuitively, that suggests that the equilibrium fraction of $L-$alleles must become smaller. Intuitively, if the initial distribution of alleles in the population is such that (22) is violated, we expect a number of H-alleles to eventually become fixed, which is equivalent to a reduction in n. The process

would continue until n is small enough for the number of relevant activities not to be too large, so that (22) holds.

REFERENCES

Saint-Paul, Gilles (2002), "On market forces and human evolution", CEPR Working paper, London.