Stability and Explanatory Significance of Some Simple Evolutionary Models

Brian Skyrms

University of California, Irvine

1. Introduction. The explanatory value of equilibrium depends on the underlying dynamics. First there are questions of dynamical stability of the equilibrium that are internal to the dynamical system in question. Is the equilibrium locally *stable*, so that states near to it stay near to it, or better, *asymptotically stable*, so that states near to it are carried to it by the dynamics? If not, we should not expect to see this equilibrium. But even if an equilibrium is asymptotically stable, that is no guarantee that the system will reach that equilibrium unless we know that the system's initial state is sufficiently close to the equilibrium. Global stability of an equilibrium, when we have it, gives the equilibrium a much more powerful explanatory role. An equilibrium is *globally asymptotically stable* if the dynamics carries every possible initial state in the interior of the state space to that equilibrium. If an equilibrium is globally stable, it can have explanatory value even when we are completely uncertain about the initial state of the system.

Once questions of dynamical stability are answered with respect to the dynamical system in question, there is the further question of structural stability of that system itself. That is to say, are dynamical systems close to the one in question (in a sense to be made

precise) topologically equivalent to that system? If not, a slight mispecification of the model may make predictions that are drastically wrong.

Structural stability is defined in terms of small changes in the model. But we may also be interested in what happens with some rather large changes in the model. A structurally stable model might, after all, be badly mispecified. Interest in such questions depends on the plausibility of the large changes being contemplated.

Here I would like to discuss these stability questions with respect to three simple dynamical evolutionary models from my book, *Evolution of the Social Contract.* Two are models of simplified bargaining games, one with random encounters and one with correlated encounters. The third is a model of a simplified signaling game. These models all use replicator dynamics.

In modeling evolution - even cultural evolution - we face considerable uncertainty concerning early states of the system. Considerations of dynamical stability are therefore crucial to the evaluation of the explanatory significance of the model. If we can only show that an equilibrium is stable (that is to say, *locally* stable), we have only a "how possible" explanation. If we can show global stability, then the early state of the system is immaterial and, providing the dynamical system is the correct model, we have approach a "why necessarily" explanation.

Good reasons can be given why the replicator dynamics is a plausible candidate for a dynamics of cultural evolution. A number of different models of social learning by imitation have been shown to yield the replicator dynamics. [Binmore, Gale and Samuelson (1995), Björnerstedt and Weibull (1995), Sacco (1995), Schlag (1998)]. As I have said elsewhere, I believe that the replicator dynamics is a natural place to begin investigations of dynamical models of cultural evolution, but I do not believe that it is the whole story. [Skyrms (1999)] That means that structural stability, and more generally, stability under perturbations of the dynamics itself, are important. The more robust the result is to perturbations of the dynamics, the more likely it is of real significance in cultural evolution.

In *Evolution of the Social Contract* I make, without presenting proof, some claims about stability. I will substantiate those claims here. I will prove local asymptotic dynamic stability in the bargaining game with random encounters and global asymptotic dynamical stability in the other two models.

The dynamical stability results are of a rather different character for the bargaining game of chapter one than for the signaling game of chapter five. Accordingly I make explanatory claims of different strengths for these two cases. In the bargaining game with random encounters, there are two attracting equilibria: one where everyone settles for equal shares and one where there is a population in which some demand a lot and some demand a little. I argue that, in the game in question, the equal division equilibrium in the one selected by commonly held norms of justice. The equal division

equilibrium has the largest basin of attraction, but the basin of attraction of the inegalitarian equilibrium is not so small as to be negligible. At this point the power of the dynamics to explain cultural evolution of egalitarian norms is not impressive, and I say so. However, it may be that such norms evolved in an environment where encounters are not random, but there is positive correlation between types - for which various reasons might be given. A small amount of positive correlation incorporated in the second bargaining model eliminates the inegalitarian basin of attraction, and makes the equal division equilibrium a global attractor. (Larger amounts of positive correlation have the same qualitative result.) I conclude: "This is, perhaps, the beginning of an explanation of the origin of our concept of justice." [Skyrms (1996) p. 21]. The words were chosen carefully. The claim of explanatory significance is meant to be modest.

In the case of the signaling game, the dynamical stability properties are quite different. In the model with random encounters and with one signaling system equilibrium and two anti-signaling system equilibria (my third model), the signaling system equilibrium is a global attractor. This remains true if positive correlation is added. In larger models with many signaling system equilibria, almost all possible initial populations are carried to one signaling system equilibrium or another. In this case the dynamical stability results are much stronger than in the bargaining game, and I make a much stronger explanatory claim: "The emergence of meaning is a moral certainty." [Skyrms (1996) p. 93].

In a recent article, D'Arms, Batterman and Górny, with whom I agree on the importance of stability, raise questions about the structural stability of my second model. [D'Arms, Batterman and Górny (1998) p. 91.] In fact, I will show that this is the *only* one of the three models that *is* structurally stable. Modification of model 3 to allow correlated encounters, however, results in a structurally stable system. I will then show that a number of the foregoing results remain true if we substitute any dynamics in a large class of *qualitatively adaptive* dynamics. This strengthens the explanatory force of my models, because it identifies behavior that does not depend on the replicator dynamics or some dynamics very close to it being the right dynamics to use in a theory of cultural evolution. Of course, it is possible that dynamics that are even further afield may be of interest and importance. But I think that it is useful to have the questions addressed here settled. The techniques used to do so may also be of interest in the analysis in more complicated models.

2. **The Three Dynamical Models**

The dynamics underlying each of these models is the *replicator dynamics*. The proportion of the population playing strategy *i* is x_i. In each of the models considered here there are three strategies, so the state of the population is a vector, $\mathbf{x} = \langle x_1, x_2, x_3 \rangle$. The state space is the 3-simplex where $x_1 + x_2 + x_3 = 1$, with x_1, x_2, x_3, non-negative, as shown in figure 1.

(Figure 1 here)

The average fitness of strategy *i* in population state **x** is denoted by "U(x_i | **x**)". [For future reference we also U(**y** | **x**), the average fitness of a in infinitesimal subpopulation playing strategies in the proportion specified by vector **y** when random paired with members of a population in states **x**, as $_i\, y_i \bullet U(x_i \mid \mathbf{x})$.] The average fitness of the whole population in state **x** is denoted by "U(**x** | **x**)". The *replicator dynamics* is then given by the system of differential equations:

$$dx_i/dt = x_i \quad [U(x_i \mid \mathbf{x}) - U(\mathbf{x} \mid \mathbf{x})]$$

The 3-simplex is invariant under this dynamics. If **x** is in the 3-simplex at a time, it remains within the 3-simplex for all time. The dynamics for three strategies thus lives on a plane.

The replicator dynamics was introduced by Taylor and Jonker (1978) as a simplified model of differential reproduction underlying the notion of evolutionarily stable strategy introduced by Maynard Smith and Price (1973). It has also been derived as the dynamics of various models of cultural evolution [Binmore, Gale and Samuelson (1995), Björnerstedt and Weibull (1996), Schlag (1998)] and as a limiting case of reinforcement learning [Borgers and Sarin (1997)].

Model 1.

The three models differ in how fitnesses are determined. In the first model individuals are paired at random from an infinite population to play a bargaining game. The three strategies are S_1: Demand 1/3 of the cake, S_2:Demand 2/3 of the cake, and S_3:

Demand 1/2 of the cake. If demands total more that 1, no one gets anything; otherwise, players get what they demand. Fitness here equals amount of cake. Those who demand 1/3 have demands that are compatible with all players and always get 1/3:

$$U(x_1 \mid \mathbf{x}) = 1/3$$

Those who demand 2/3 only get their demand when they are paired with those who demand 1/3, otherwise they get nothing. Their average fitness is:

$$U(x_2 \mid \mathbf{x}) = 2/3 \quad x_1$$

Those who demand 1/2 get their demand except when matched against those who demand 2/3. Their average fitness is:

$$U(x_3 \mid \mathbf{x}) = 1/2 \quad (1\text{-}x_2)$$

The average fitness of the population is gotten by averaging the average fitnesses of the various strategies:

$$U(\mathbf{x} \mid \mathbf{x}) = x_1 \quad U(x_1 \mid \mathbf{x}) + x_2 \quad U(x_2 \mid \mathbf{x}) + x_3 \quad U(x_3 \mid \mathbf{x})$$

Model 2.

In the second model the strategies and basic game are the same as in model one, but the individuals are not paired at random. Rather, there is some positive correlation in the encounters determined by a parameter, e. The probability that strategy i meets itself, $p(S_i|S_i)$ is not simply the proportion of the population playing that strategy, x_i, as in the random pairing model, but rather it is inflated thus:

$$p(S_i|S_i) = x_i + e \bullet (1\text{-}x_i)$$

The probability of strategy S_i meeting a different strategy, S_j is correspondingly deflated:

$$p(S_j|S_i) = x_j - e \bullet x_j$$

We will take e = 1/5, which is a case discussed both in chapter 1 of my book and in D'Arms, Batterman, and Górny. Then, as before:

$$U(x_1 \mid \mathbf{x}) = 1/3$$

but :

$$U(x_2 \mid \mathbf{x}) = 2/3 \quad 4/5 \quad x_1$$

and:

$$U(x_3 \mid \mathbf{x}) = 1/2 \quad (x_3 + 1/5\ (1-x_3)) + 1/2 \quad 4/5 \quad x_1$$

The average fitness of the population is calculated as before.

Model 3.

This is a random pairing model like model 1, but the underlying game is a signaling game from Chapter 5, pp. 91-93 of my book. There are two antisignaling system strategies, x_1, x_2, and one signaling system strategy, x_3. Each antisignaling system strategy does badly against itself for a fitness of zero, well against the other for a fitness of one and middling against the signaling system for fitness of one-half. The signaling system has fitness one-half against either of the antisignaling systems and fitness of one against itself. This gives the following average fitnesses for the three strategies:

$$U(x_1 \mid \mathbf{x}) = x_2 + (1/2) \quad x_3$$

$$U(x_2 \mid \mathbf{x}) = x_1 + (1/2) \quad x_3$$

$$U(x_3 \mid \mathbf{x}) = x_3 + \ 1/2 \quad (x_1 + x_2)$$

3. **Local Dynamical Stability of Equilibria**

First we will identify the equilibria and their local stability characteristics in each of our three models. The equilibria can be identified by solving equations. Their dynamical stability characteristics can be investigated by evaluation the eigenvalues of the Jacobian matrix of partial derivatives at the equilibrium point. [See, for instance, Hirsch and Smale (1974) Ch. 9.] If the eigenvalues all have non-zero real part, the equilibrium is said to be *hyperbolic* and the eigenvalues determine the local dynamical stability properties of the equilibrium. If the real parts of the eigenvalues are all negative, then the equilibrium is called a *sink*, and it is asymptotically stable. If the real parts of the eigenvalues are a positive, the equilibrium is called a *source*, and if both positive and negative real parts occur it is called a *saddle*. Sources and saddles are unstable. If the point is non-hyperbolic, local stability must be investigated by different means.

Model 1.

Recall that this is the bargaining game with just three strategies, S_1 (Modest) = Demand 1/3; S_2 (Greedy) = Demand 2/3; S_3(Fair) = Demand 1/2. Each vertex of the simplex represents a state in which the population is totally composed of one of the types. Here, as always in the replicator dynamics, each vertex is a dynamic equilibrium. (If the other guys are extinct, they can't reproduce.) At an equilibrium in the interior of the S_1-S_2 edge, both these strategies must have the same fitness. Solving the equations: $x_1 + x_2 = 1$ and $(1/3) = (2/3)\, x_1$, we find a Modest-Greedy equilibrium at $x_1=1/2$, $x_2=1/2$. There are no equilibria interior to the other edges. An equilibrium interior to the simplex must have all three strategies having equal fitnesses. Solving these equations we find a

Modest-Greedy-Fair equilibrium at x_1=1/2, x_2=1/3, x_3=1/6. These five states are the only equilibrium states in this model.

We proceed to examine the eigenvalues of the Jacobian at these points. Since the state space - the three-simplex- is a two-dimensional object, we need only consider the dynamics in terms of two of the three variables, any two will do. For each point we will get two eigenvalues. (We could evaluate eigenvalues for the Jacobian in the three variable system, but this would generate one spurious zero eigenvalue, associated with the eigenvector pointing out of the simplex. See Bomze(1986) p.48 or van Damme(1987) p. 222.) All the calculations of eigenvalues reported in this paper were performed using *Mathematica*. The calculations for Model 1 are given in the appendix.

At the All Fair equilibrium, x_3 = 1 the eigenvalues are {-1/2, -1/6}. The two negative eigenvalues identify this as a *sink* - an *asymptotically stable* equilibrium. The Modest-Greedy equilibrium at x_1=x_2=1/2 is likewise *asymptotically stable* with a pair of negative eigenvalues {-1/6, -1/12}. The Modest-Greedy-Fair equilibrium in the interior of the simplex at x_1=1/2, x_2=1/3, x_3=1/6, has one negative and one positive eigenvalue. The values are approximately {-0.146525, 0.061921}. This indicates that this equilibrium is a *saddle*, which is attracting in one direction and repelling in another. This equilibrium is dynamically unstable. The All Modest equilibrium at x_1=1 has two positive eigenvalues{1/6,1/3}, which identifies it as a *source*, an unstable repelling equilibrium. This leaves All Greedy equilibrium at x_2= 1. This has eigenvalues {0,1/3}. Because of the zero eigenvalue, this is not a hyperbolic equilibrium, and its local dynamical stability

properties cannot be completely inferred from the eigenvalues of the Jacobian. The one positive eigenvalue, however, shows that it is not stable. We will be able to say more about it, using different techniques, in the next section. The information that we have about model 1 so far is summarized in Table 1.

EQUILIBRIUM	EIGENVALUES	STABILITY
x_1=1, x_2=0, x_3=0	1/6, 1/3	Unstable (source)
x_1=0, x_2=1, x_3=0	0, 1/3	(non-hyperbolic)
x_1=0, x_2=0, x_3=1 (All Fair)	-1/2, -1/6	Stable (sink)
x_1=1/2, x_2=1/2, x_3=0	-1/6, -1/12	Stable (sink)
x_1=1/2, x_2=1/3, x_3=1/6	-0.146525, 0.0631921	Unstable (saddle)

Table 1

Model 2.

Model 2 is the bargaining game with correlation of encounters at the .2 level. There are four equilibria. Each of the vertices is an equilibrium. There is an equilibrium on the Greedy-Modest edge, at x_1=5/8, x_2=3/8. There is no equilibrium in the interior of the 3-simplex. When we evaluate the eigenvalues of the Jacobian at these equilibria we find that they are all hyperbolic, so we have a complete characterization of the local stability characteristics of these equilibria. These are collected in the following table:

EQUILIBRIUM	EIGENVALUES	STABILITY
x_1=1, x_2=0, x_3=0	1/6,1/5	Unstable (source)
x_1=0, x_2=1, x_3=0	1/10,1/3	Unstable (source)
x_1=0, x_2=0, x_3=1 (All Fair)	-1/2, -1/6	Stable (sink)
x_1=5/8, x_2=3/8, x_3=0	-1/8, 1/60	Unstable (saddle)

Table 2

Model 3.

In model 3, x_1 and x_2 are two "antisignaling system strategies" and x_3 is a signaling system strategy. There are four equilibria - the three vertices and an antisignaling polymorphism at $x_1 = x_2 = 1/2$. We have three hyperbolic equilibria and one non-hyperbolic equilibrium. The analysis is summarized in table 3.

EQUILIBRIUM	EIGENVALUES	STABILITY
x_1=1, x_2=0, x_3=0	1/2, 1	Unstable (source)
x_1=0, x_2=1, x_3=0	1/2, 1	Unstable (source)
x_1=0, x_2=0, x_3=1 (Signaling)	-1/2, -1/2	Stable (sink)
x_1=1/2, x_2=1/2, x_3=0	-1/2, 0	(non-hyperbolic)

Table 3

4. **Global Dynamical Stability of Equilibria**

The sinks identified in section 3 are dynamically asymptotically stable. This is a local property that is established by dynamical behavior in some neighborhood of the equilibrium. But the neighborhood might be very small. It would be more powerful if we could demonstrate that an equilibrium has a significant basin of attraction, or even that it is globally asymptotically stable - that the dynamics carries every point in the interior of the state space to it, in the limit.

The main technique employed in this section is the use of the Kullbach-Leibler relative entropy as a Liapunov Function for replicator dynamics. A Liapunov function is a generalization of the notion of potential. Liapunov showed that if **x** is an equilibrium and V is a continuous real-valued function defined on some neighborhood, W, of **x,** differentiable on W-**x** such that :

(i) V(**x**)=0 and for **y** **x** in W, V(**y**)>0.

(ii) The time derivative of V, V' < 0 in W-**x**.

Then the orbit of any point in W approaches **x** as time goes to infinity. If these requirements can be shown to hold taking the neighborhood as the whole state space (or its interior), then **x** is *globally asymptotically stable*. [See Hirsch and Smale ,Ch.9 Sec 3. and Guckenheimer and Holmes pp. 5ff.].

The Kullback-Leibler relative entropy serves as an appropriate Liapunov function. The Kullback-Leibler relative entropy of state **y** with respect to state **x** is:

$$H_{\mathbf{x}}(\mathbf{y}) = \sum_i x_i \log (x_i / y_i)$$

with the sum being taken over the carrier of **x**, that is to say over the strategies that have positive population proportion in state **x**. This function meets the requirements of continuity and differentiability for a Liapunov function and assumes its minimum at **x**.

Its time derivative of $H_{\mathbf{x}}(\mathbf{y})$ is just -[U(**x** | **y**) - U(**y** | **y**)], which we need to be negative to complete the requirements for a Liapunov function. So we only need to check that [U(**x** | **y**) - U(**y** | **y**)] is positive throughout the neighborhood in question. . [See Bomze (1991) and Weibull(1997) 3.5 and 6.5.]

Model 3

We can now prove global convergence to the signaling system equilibrium in the signaling game of model 3. The state of fixation of the signaling system strategy is at x_3=1. We consider the neighborhood, $x_3 > 0$, where this strategy is not extinct This includes the vertex, x_3=1, the x_1-x_3 and x_2-x_3 edges, and the interior of the 3-simplex. We use as a Liapunov function the entropy relative to the state where x_3=1. We now need to check that if x_3>0, U(x_3 | **x**) - U(**x** | **x**) is non-negative everywhere and equal to zero only where $x_3 = 1$.

First, we want to show that for a fixed value of x_3, U(x_3 | **x**) - U(**x** | **x**) assumes its minimum at the point where x_1=x_2. Then we need only check that on the line, x_1=x_2, to establish the desired result. For a fixed values of x_3, U(x_3 | **x**) = 1/2 x_1 + 1/2 x_2 + x_3 is constant so U(x_3 | **x**) - U(**x** | **x**) assumes its minimum when U(**x** | **x**) assumes its maximum.

For fixed x_3, $U(\mathbf{x} \mid \mathbf{x}) = x_1 U(x_1) + x_2 U(x_2) + x_3 U(x_3)$ assumes its maximum when $x_1 U(x_1) + x_2 U(x_2)$ does. The quantity, $x_1 U(x_1) + x_2 U(x_2) = x_1(x_2 + x_3/2) + x_2(x_1+x_3/2) = 2x_1x_2 + x_3/2(x_1 + x_2)$ now assumes its maximum where $2x_1x_2$ does, that is where $x_1=x_2$.

Now consider a point, **x**, on the line, $x_1 = x_2$. Here $U(x_3|\mathbf{x}) = (1/2)(2x_1) + (1- 2x_1) = 1-x_1$. And $U(x_2|\mathbf{x}) = U(x_1|\mathbf{x}) = x_2 + (1/2)\, x_3 = x_1 + (1/2)(1-2\, x_1) = 1/2$. So when x_1, x_2, x_3, are all positive, $U(x_3 \mid \mathbf{x}) > U(\mathbf{x} \mid \mathbf{x})$ when $(1-x_1) > 1/2$. On the line $x_1 = x_2$ in the interior of the simplex, $1/2 > x_1 > 0$ so , $U(x_3 \mid \mathbf{x}) > U(\mathbf{x} \mid \mathbf{x})$. By definition at $x_3=1$, $U(x_3 \mid \mathbf{x}) = U(\mathbf{x} \mid \mathbf{x})$. Thus, in the global neighborhood, $x_3 > 0$, where the signaling strategy is not extinct, our Liapunov function is non-negative and equal to zero only at the point of fixation of that strategy. In Model 3 we have global convergence to the signaling system equilibrium.

This result answers the lingering question about the local stability properties of the equilibrium $x_1= x_2=1/2$, that was left by zero the eigenvalue that showed up in the local stability analysis of the previous section. This equilibrium is locally dynamically unstable within the 3-simplex. However, within the subsimplex consisting of the x_1-x_2 line (where x_3 is extinct) this equilibrium is globally stable. This can be shown by applying the same techniques within the sub-simplex.

Model 2.

The qualitative analysis of Model 2 is much like that of model 3. We use the same relative entropy Liapunov function. The quantity $U(x_3 \mid \mathbf{x}) - U(\mathbf{x} \mid \mathbf{x})$, which simplifies to

1/30 (x_1 (5-28 x_2) + 3(5-4 x_2) x_2), is positive throughout the entire region where x_3>0. Here I will let a picture stand in for the algebraic proof. Figure 2 is a plot of this function above the x_1-x_2 plane. It is positive on the relevant region [x_1>0, x_2 >0, ~(x_1 + x_2 >1)]. The line x_3 = 0, is a subsimplex. On this subsimplex the Greedy-Modest mixed equilibrium is global asymptotically stable (even though it is not even locally stable on the whole 3-simplex).

Model 1.

In model 1, there is no globally asymptotically stable equilibrium. Both the All Fair equilibrium, *x3* = 1, and the Greedy-Modest equilibrium, $x_2 = x_3 = 1/2$, are asymptotically stable sinks. They both have basins of attraction that include points in the interior of the 3-simplex. Liapunov functions can be used as before, however, to prove that an equilibrium attracts all points in some significantly extended neighborhood. As an illustration, consider the neighborhood of the All Fair equilibrium defined by $x_3 > .7$. (This is by no means the whole basin of attraction for this equilibrium, but it makes for a quick example.) It is easy to see that U(x_3 | **x**) - U(**x** | **x**) must be positive throughout this neighborhood. The minimum value that U(x_3 | **x**) can have is (.7)(.5) = .35. U(x_1 | **x**) is constant at 1/3. U(x_2 | **x**) has its maximum at (2/3)(.3) = .2. So U(x_3 | **x**) must always be greater than the average, U(**x** | **x**). We will have reason refer back to this illustration in section 6.

To summarize, we have established:

4.1 the All Fair and Signaling System equilibria, of models 2 and 3 respectively, are *globally asymptotically stable*.

4.2 The All Fair equilibrium in model 1 is *locally asymptotically stable* and attracts every point in the neighborhood $x_3 > .7$.

5. **Structural Stability of the Dynamical System**

A dynamical system system is *structurally stable* if small enough, but otherwise arbitrary, perturbations result in a topologically equivalent system. That is to say that all dynamical systems sufficiently close to the one under consideration, are topologically equivalent. This will be made precise shortly.

It is sometimes held that only structurally stable dynamical systems have explanatory value, but this *stability dogma*, has also been questioned (as in Guckenheimer and Holmes, Ch. 5.) The point is that the only perturbations that are crucial are those that are physically possibly for the phenomena under consideration. To demand structural stability may be to demand too much. It might also be to demand too little, as plausible perturbations of the dynamical system might be large.

Nevertheless, is may be useful to see whether a dynamical system is structurally stable, and if it fails to be so to see how it fails. With regard to the models under consideration here, the question of structural stability of model 2 is raised by D'Arms,

Batterman and Górny (1998) p. 91. (They also have concerns about larger perturbations of the model.) It may somewhat surprising, then, that we will be able to prove that of our three models only Model 2 is structurally stable.

First, we need a precise definition of structural stability. [Piexoto (1962), Smale (1980), Guckenheimer, J. and Holmes, P. (1986)]. In each of our models, the differential equations of the replicator dynamics generate a dynamical system on the compact subset of the x_1-x_2 plane, where x_1 and x_2 are non-negative and their sum does not exceed one. Call this region, M. (As noted earlier, we need only consider this system, since the values of x_1 and x_2 determine the value of x_3.) The planar nature of M simplifies the following discussion in a number of ways.

The differential equations define a continuously differentiable *vector field* on this region, M. The vector field is a function that associates with every point the vector $\langle dx_1/dt, dx_2/dt \rangle$ with the derivatives being evaluated at that point. This vector field can be thought of as being the dynamical system. It determines a family of *solution curves*, or *orbits*, on M. Two dynamical systems are *topologically equivalent* if there is a homeomorphism from M to M taking which preserves orbits and their temporal sense.

To consider dynamical systems "close" to one of our models, we need a space of dynamical systems and a sense of closeness. For the space of dynamical systems, we take the continuously differentiable vector fields on M, (M). Closeness can be defined relatively simply because of the nice nature of our underlying state space, M. For each

dynamical system, X in (M), let its norm, ||X||, be that maximum of the following numbers:

least upper bound | dx_1/dt |

least upper bound | dx_1/dt |

least upper bound | $[dx_1/dt]/ x_1$ |

least upper bound | $[dx_1/dt]/ x_2$ |

least upper bound | $[dx_2/dt]/ x_1$ |

least upper bound | $[dx_2/dt]/ x_2$ |

with the least upper bounds being taken over all points in M. Now we can define a *metric* on the space of dynamical systems, (M). For two such dynamical systems, X, Y, we take $d(x,y) = ||X - Y||$. Now we can give a precise definition of structural stability:

A vector field, X, is *structurally stable*, if there is a neighborhood of X such that every vector field, Y, in that neighborhood is *topologically equivalent* to X.

The tool to be used in this section to establish structural stability is Peixoto's Theorem [Peixoto (1962), Guckenheimer and Holmes(1986) p. 60.]:

A vector field defined on a compact region of the plane is structurally stable ***if an only if***

(1) The number of equilibrium points and closed orbits is finite and each is hyperbolic.

(2) There are no orbits connecting saddle points.

(We note again that the planar nature of M simplifies the form of the theorem used here.)

Since the conditions are necessary for structural stability, and since models 1 and 3 have non-hyperbolic equilibria, we can conclude that these dynamical systems are not structurally stable. What about model 2? In section 3, we saw that it has a finite number of equilibria, all of which are hyperbolic. In section 4 we saw that the "All Fair" equilibrium at x_3=1 is globally asymptotically stable for the subregion where x_3>0. The leaves only the line x_3=0. We saw that the Greedy-Modest mixed equilibrium is globally asymptotically stable on the interior of that line. Therefore there are no closed orbits and there are no orbits connecting saddles. Since the conditions are sufficient for structural stability, Model 2 is structurally stable.

5.1 Models 1 and 3 are not structurally stable

5.2 Model 2 is structurally stable.

(Analysis of structural stability is more complicated in higher dimensional systems. Model 2 is a Morse-Smale system [Guckenheimer and Holmes p. 64.] In higher dimensions being Morse-Smale is a sufficient condition for structural stability, but is no longer a necessary condition.)

6. **Qualitatively Adaptive Dynamics**

Let us say that a dynamics is *qualitatively adaptive* if, according to that dynamics, a strategy that is not extinct increases its proportion if the population if its fitness is higher that the population average, decreases its population proportion if its fitness is less that the population average, and keeps the same population proportion if its fitness is equal to the population average. That is to say:

If $x_i > 0$, then dx_i/dt agrees in sign with $U(x_i \mid \mathbf{x}) - U(\mathbf{x} \mid \mathbf{x})$

The replicator dynamics is a member of this class, but there are many other qualitatively adaptive dynamics. Substituting another qualitatively adaptive dynamics for the replicator dynamics in out three models can significantly perturb the vector field. Structural stability does not guarantee anything about behavior of the dynamical systems under this class of perturbations. Nevertheless we can show that the global stability results of section 4 generalize to all these systems.

We use the same relative entropy as a Liapunov function. To prove convergence to the equilibrium as x_3=1, the entropy relative to the equilibrium at state **x** is just:

$$- \log (x_3)$$

This function is continuous and non-negative on the simplex and equals zero only at the equilibrium. Its time derivative is:

$$-(1/x_3)\, dx_3/dt$$

To prove convergence, we now need only to show that this is negative on the region consisting of the neighborhood x_3>0 with the equilibrium removed - that its to say when $0 < x_3 < 1$. Here the time derivative is negative just where dx_3/dt is positive, which is where $U(x_3 \mid \mathbf{x}) - U(\mathbf{x} \mid \mathbf{x})$ is positive since the dynamics is adaptive.

Thus the results of section 4 carry over:

6.1 With any *qualitatively adaptive dynamics* substituted for the replicator dynamics, the All Fair and Signaling System equilibria, of models 2 and 3 respectively, are *globally asymptotically stable.*

6.2 For any *qualitatively adaptive dynamics* substituted for the replicator dynamics, the All Fair equilibrium in model 1 is *asymptotically stable* and attracts every point in the neighborhood $x_3 > .7$.

We can even strengthen **6.2** to include a broader class of dynamics. Say that a dynamics is *weakly qualitatively adaptive* if a strategy, s_i, has positive growth in population proportion, $dx_i/dt > 0$, in any state in which it is not extinct and its fitness is the highest of any strategy represented in the population. (A dynamics may be weakly qualitatively adaptive, without being qualitatively adaptive if, for instance, being second best and better than the average counts for nothing.) Now we say that in model 1, if $x_3 > .7$, then strategy 3, Demand 1/2, is the fittest strategy. Then by the same reasoning as before we have:

6.3 For any *weakly qualitatively adaptive dynamics* substituted for the replicator dynamics, the All Fair equilibrium in model 1 is *asymptotically stable* and attracts every point in the neighborhood $x_3 > .7$.

In chapter 1 of *Evolution of the Social Contract*, (p.11), in a passage quoted by D'Arms, Batterman and Gorny, I claimed that the fact that the All Fair equilibrium is asymptotically stable was robust for a wide class of adaptive dynamics:

> [Demand 1/2's] strong stability properties guarantee that it is an attracting equilibrium in the replicator dynamics, but also make the details of that dynamics

unimportant. Fair division will be stable in any dynamics with a tendency to increase the proportion (or probability) of strategies with greater payoffs, because any unilateral deviation from fair division result in a strictly worse payoff. For this reason, the Darwinian story can be transposed into the context of *cultural evolution*, in which imitation and learning play an important role in the dynamics. Proposition **6.2** establishes the technical claim that I make in this passage, and proposition **6.3** strengthens it in a way that supports my general conclusion.

7. **Correlation Structure**

Is model 2 robust against arbitrary changes in the correlation structure? Model 1 shows that it is not. Model 2 is not topologically equivalent to model 1. If you take model 2 and gradually reduce the correlation you come to a point where there is a bifurcation and the four equilibrium points of model 2 become the five equilibrium points of model 1. The modest-greedy equilibrium changes from a saddle to a sink. Arbitrary changes to the correlation structure can make a qualitative difference here, as they can in almost any game. (D'Arms, Batterman and Górny (1998) make the same point by constructing an alternative model with anti-correlation.) But we know that model 2 is robust against local changes in correlation structure, those within some small neighborhood, as a result of our structural stability result.

Model 3 is not structurally stable. What happens to it if we add a small amount of positive correlation? We get:

Model 4

$$U(x_1 \mid \mathbf{x}) = (1\text{-}e)\quad x_2 + (1/2)\quad (1\text{-}e)\quad x_3$$

$$U(x_2 \mid \mathbf{x}) = (1\text{-}e)\quad x_1 + (1/2)\quad (1\text{-}e)\quad x_3$$

$$U(x_3 \mid \mathbf{x}) = x_3 + e\quad (x_1 + x_2) + 1/2\quad (1\text{-}e)\quad (x_1 + x_2)$$

With e=0, Model 4 reduces to model 3 and the equilibrium at $x_1 = x_2 = 1/2$ is non-hyperbolic with eigenvalues {-1/2, 0}. In model 4, the eigenvalues of the Jacobian evaluated at this equilibrium are:

$$\{1/4\ (-1 + 3e - \mathrm{SQR}(1 + 2e + e^2),\ 1/4(-1 + 3e + \mathrm{SQR}(1 + 2e + e^2)\}$$

For any small positive e, the equilibrium becomes hyperbolic. For example, for e=.001, the eigenvalues are {-0.4995, 0.001}. The equilibrium is a saddle, and thus unstable. The local stability properties of the other equilibria remain unchanged. We now have a *structurally* stable system that bears a qualitative resemblance to model 2. [That is to say that there are two sources at x_1=1 and at x_2=1, a mixed equilibrium saddle on the x_1-x_2 line, and a sink at x_3=1.]

Clearly, correlation structure can be very important in determining the stability properties of an evolutionary dynamical system. It is of particular interest to consider alternative models where correlation is endogenously generated. D'Arms, Batterman, and Górny have ideas in this direction that might be profitably pursued in more detail. If I am not mistaken, with easier (and therefore higher levels of) correlation and anti-correlation of the kind that they consider in their paper, the Greedy-Modest equilibrium again becomes unstable. It would be interesting to have the whole story.

An alternative route to endogenous correlation structure uses interactions with neighbors in a spatially explicit structure. In Alexander (1999) and Alexander and Skyrms (1999) it is shown that the correlation that emerges in this sort of model for bargaining games (like that of model one) has dramatic effects on both the proportion of initial conditions that evolve to the All Fair equilibrium, and the speed with which that process takes place.

8. **Stability and Explanatory Significance**

Stability considerations are only one ingredient to be taken into account in assessing the explanatory significance of dynamical models. There are also larger considerations regarding the aptness of the model (or class of models) to the process being modeled - in this case the process of cultural evolution. I will not address these larger considerations here. They have been discussed elsewhere. [D' Arms (forthcoming), Barrett, Eells, Fitelson and Sober (1999), Bicchieri (1999), Bolton(1997), (forthcoming), Carpenter (forthcoming), Gintis (forthcoming), Güth and Güth (forthcoming), Harms (forthcoming), Kitcher (1999), Krebs (forthcoming), Mar(forthcoming), Nesse (forthcoming), Proulx(forthcoming), Skyrms (1999), (forthcoming)]. But, bracketing these concerns, we can ask what bearing the stability results of this paper have on the explanatory significance of the models discussed here.

The fact that the All Fair equilibrium is an attractor in the bargaining game of Models 1 and 2 is extremely robust. It holds not only in both these models, but also in models in which the replicator dynamics is replaced by any other dynamics in the class of weakly qualitatively adaptive dynamics. But the All Fair equilibrium is not a global attractor in model one. Adding a modest amount of positive correlation (e=1/5 or more) turns it into one, in a structurally stable dynamical system. Independent arguments for this degree of positive correlation during the evolution of the norm (or for some modified dynamical model) would be required to complete the explanation. My models do not give the whole story of the evolution of the equal split - they begin that story.

The fact that the signaling system equilibrium is a global attractor in Model 3 and remains so if the replicator dynamics is replaced by any qualitatively adaptive dynamics, strikes me as explanatorily powerful. But the fact that model 3 is not structurally stable demands attention. Adding positive correlation takes us to model 4, which is structurally stable, and in which the signaling system equilibrium remains a global attractor. Here - in contrast to the case of the bargaining game - any amount of positive correlation, no matter how small, will do.

In general, these results should focus our attention on correlation. Correlation structure is of the utmost importance. Models of the co-evolution of strategic interaction with correlation structure are a promising direction for future research.

REFERENCES

Alexander, J. (1999) "The (Spatial) Evolution of the Equal Split" Institute for Mathematical Behavioral Science. U. C. Irvine.

Alexander, J. and Skyrms, B. (1999) "Bargaining with Neighbors: Is Justice Contagious?" *Journal of Philosophy* XCVI , 588-598.

Andronov, A. A. and Others (1971), *Theory of Bifurcations of Dynamical Systems on a Plane.* (tr. from the Russian Original of 1967) Israel Program of Scientific Translations: Jerusalem.

D'Arms, J. (forthcoming) "When Evolutionary Game Theory Explains Morality, What Does it Explain?" *Journal of Consciousness Studies*.

D'Arms, J. (1996) "Sex, Fairness and the Theory of Games" *Journal of Philosophy* 96: 615-127.

D'Arms, J., Batterman, R. and Górny, K. (1998), "Game Theoretic Explanations and the Evolution of Justice", *Philosophy of Science* 65: 76-102.

Barrett, M., Eells, E. Fitelson, B. and Sober, E. (1999) "Models and Reality: A Review of Brian Skyrms's *Evolution of the Social Contract*" *Philosophy and Phenomenological Research* 59: 237-241.

Bicchieri, C. (1999) "Local Fairness" *Philosophy and Phenomenological Research* 59: 229-236.

Binmore, K., Gale, J. and Samuelson, L. (1995) "Learning to be Imperfect: The Ultimatum Game" *Games and Economic Behavior* 8: 56-90.

Björnerstedt, J. and Weibull, J. (1996) "Nash Equilibrium and Evolution by Imitation" In K. Arrow et. al. (eds.) *The Rational Foundations of Economic Behavior* Macmillan: New York, 155-171.

Bolton, G. (1997) "The Rationality of Splitting Equally" *Journal of Economic Behavior and Organization* 32:365-381.

Bolton, G. (forthcoming) "Motivation and the Games People Play" *Journal of Consciousness Studies*.

Bomze, I. M. (1991) "Cross-Entropy Minimization in Uninvadable States of Complex Populations" *Journal of Mathematical Biology* 30:73-87.

Bomze, I. M. (1986) "Non-Cooperative Two Person Games in Biology: A Classification" *International Journal of Game Theory* 15:31-59.

Borgers, T. and Sarin, R. (1997) "Learning Through Reinforcement and the Replicator Dynamics" *Journal of Economic Theory* 77:1-14.

Carpenter, J. (forthcoming) "Blurring the Line Between Rationality and Evolution" *Journal of Consciousness Studies*.

van Damme, E. (1987) *Stability and Perfection of Nash Equilibria* Berlin: Springer.

Gintis, H. (forthcoming) "Classical vs. Evolutionary Game Theory" *Journal of Consciousness Studies*.

Guckenheimer, J. and Holmes, P. (1986), *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields.* New York: Springer.

Güth, S. and Güth, W. (forthcoming) "Rational Deliberation versus Behavioral Adaptation: Theoretical Perspectives and Experimental Evidence" *Journal of Consciousness Studies*.

Harms, W. (forthcoming) "The Evolution of Cooperation in Hostile Environments" *Journal of Consciousness Studies*.

Hirsch, M. and Smale, S. (1974) *Differential Equations, Dynamical Systems, and Linear Algebra*. Academic Press: New York.

Hofbauer, J. and Sigmund, K. (1988) *The Theory of Evolution and Dynamical Systems*. Cambridge University Press: New York.

Kitcher, P. (1999) "Games Social Animals Play: Commentary on Brian Skyrms' *Evolution on the Social Contract*" *Philosophy and Phenomenological Research* 59: 221-228.

Krebs, D. (forthcoming) "Evolutionary Games and Morality" *Journal of Consciousness Studies*.

Mar, G. (forthcoming) "Evolutionary Game Theory, Morality and Darwinism" *Journal of Consciousness Studies*.

Maynard-Smith, J. and Price, G. (1973) "The Logic of Animal Conflicts" *Nature* 246, 15-18.

Nesse, R. (forthcoming) "Strategic Subjective Commitment" *Journal of Consciousness Studies*.

Peixoto, M. M. (1962) "Structural Stability on Two-Dimensional Maniforld" *Topology* 1, 101-120.

Proulx, C, (forthcoming) "Distributive Justice and the Nash Bargaining Solution" *Journal of Consciousness Studies*.

Sacco, P. L (1995) "Comment" in K. Arrow et. al. (eds.) *The Rational Foundations of Economic Behavior* Macmillan: New York,155-71.

Schlag, K. (1998) "Why imitate, and if so how? A bounded rational approach to the multi-armed bandits." *Journal of Economic Theory* 78: 130-156.

Skyrms, B. (1994) "Darwin meets *The Logic of Decision*" *Philosophy of Science* 61:503-528

Skyrms, B. (1996), *Evolution of the Social Contract*. Cambridge University Press:New York.

Skyrms, B. (1997), "Chaos and the Explanatory Significance of Equilibrium: Strange Attractors in Evolutionary Game Dynamics" *The Dynamics of Norms* ed. C. Bicchieri et al. New York: Cambridge199-222.

Skyrms, B. (1999) "Precis of *Evolution of the Social Contract*" and "Reply to Critics" *Philosophy and Phenomenological Research* 59, 217-220 and 243-254.

Skyrms, B. (forthcoming) "Game Theory, Rationality and Evolution of the Social Contract" and "Reply to Commentary" *Journal of Consciousness Studies*.

Smale, S. (1980) *The Dynamics of Time: Essays on Dynamical Systems, Economic Processes and Related Topics.* New York:Springer.

Taylor, P. and Jonker, L. (1978) "Evolutionarily Stable Strategies and Game Dynamics" *Mathematical Biosciences* 40:145-156.

Weibull, J. (1997) *Evolutionary Game Theory* MIT Press: Cambridge, Mass.

Appendix

Calculation of Eigenvalues of the Jacobian at Equilibria for Model 1 using *Mathematica*