



**NINTH ANNUAL CONFERENCE ON
ECONOMETRIC MODELLING FOR AFRICA**

**30 June
to
2 July 2004**

Chance, Collective Action, and Institutional Innovation

Samuel Bowles

Chance, Collective Action, and Institutional Innovation

Samuel Bowles
Santa Fe Institute and University of Siena
May 17, 2004

the central problem of evolution .. is that of a trial and error mechanism by which the locus of a population may be carried across a saddle from one peak to another and perhaps higher one. This view contrasts with the conception of steady progress under natural selection... Consideration of the means by which the locus of a population may be carried across a saddle may be of interest from this standpoint.

Sewall Wright, *Journal of Genetics*, (Wright (1935):264)

Men make history; but they do not make it just as they please; they do not make it under circumstances chosen by themselves but under circumstances ... given and transmitted from the past..

Karl Marx, *The Eighteenth Brumaire of Louis Bonaparte* (1851).
(Marx (1963):15)

Introduction

The processes bringing about institutional change may involve some combination of between-group competition and within-group dynamics.¹ Here, I confine myself to the within-group processes.² Two quite distinct approaches to the within-group processes bringing about institutional innovation may be identified.

The first, similar to Sewall Wright's use of drift to explain a movement from one fitness peak across a fitness valley to another peak, is that proposed by stochastic evolutionary game theory pioneered by Foster and Young (1991) and extended by Kandori, Mailath, and Rob (1993) and others. In this Darwin-inspired approach, change occurs through the chance bunching of individuals' idiosyncratic play of non-best-responses. These will occasionally be sufficient to tip the underlying dynamic process from the basin of attraction of one conventional equilibrium to another. Changes in language use, contractual shares, market days, and etiquette have been modeled in this manner

The second approach, initiated by Marx, stresses asymmetries among the players and explains institutional innovation by the changing power balance between those who benefit from differing conventions. In this framework, revolutionary change in institutions is likely when existing institutions facilitate the collective action of those who would benefit from a change in institutions, and when, because existing institutions are inefficient by comparison to an alternative, there are substantial potential gains to making a switch. This collective-action-based approach has been used to model conflicts among classes resulting in a basic transformation of social organization, such as the French, Russian, and Cuban revolutions as well as more gradual changes in institutional arrangements such as the centuries-long erosion of European feudalism.

Do these approaches allow us to say anything about the characteristics of evolutionarily successful institutions? Though the underlying causal mechanisms are different, the Marx-inspired approach shares with Darwin-inspired stochastic evolutionary game theory the prediction that institutional arrangements which are both inefficient and highly unequal will bear an evolutionary disability and will tend to be displaced in the long run by more efficient and more egalitarian institutions.³ This is quite an arresting claim in light of the long-term historical persistence of social

¹ Thanks to the Santa Fe Institute, the MacArthur Foundation and the Pew Charitable Trusts for support of this project. This paper draws on material in chapter 12 of Bowles (2004)

² My co-authors and I develop between-group competition models of institutional evolution in a series of related papers (Bowles, Choi, and Hopfensitz (2003), Bowles and Choi (2003) and chapter 13 of Bowles (2004))

³ Efficient institutions yield a larger joint surplus, while in a more equal convention, the share of the least well-off is larger.

arrangements which would appear to be neither efficient nor egalitarian. I will explore this proposition as a way of extending the stochastic evolutionary game theoretic approach.

I begin in the next section with a simple non-stochastic population game in which the stage game exhibits two conventional equilibria. The evolution of institutions is then represented as a problem of equilibrium selection to be studied using a model of institutional persistence and accessibility. To do this, I introduce stochastic evolutionary game theory. Drawing on the work of Young and Kandori, Mailath and Rob, I show that it yields a rather strong characterization of evolutionary robust successful institutions akin to Parsons' evolutionary universals. I also give some reasons why the application of stochastic evolutionary game theory to real historical evolutions may require modulations of the model. I then augment the stochastic framework by introducing players who intentionally pursue conflicting interests through collective action. Using this extended model, I explore the long-term persistence of equal and efficient conventions when less efficient and less equal conventions are also feasible. The dynamics supported by intentional rather than accidental non-best response actions are not the same, and models incorporating intentional action in pursuit of common interests, suggest that while more efficient and more equal institutions are indeed favored by this evolutionary process under some conditions, it is also true that inefficient and unequal institutions can persist over very long periods of time.

The persistence and accessibility of historically contingent institutions

Because of their historical importance, I will focus on economic institutions that regulate the size of the social surplus and its distribution. An institution may be represented as one of a number of possible conventional equilibria in which members of a population typically act in ways that are best responses to the actions taken by others and have formed expectations that support continued adherence to these conventional actions. Examples of such distributional conventions include simple principles of division such as "finders keepers" or "first come first served," as well as more complicated principles of allocation such as the variety of rules which have governed the exchange of goods or the division of the products of one's labor over the course of human evolution. Because a convention is one of many possible mutual best responses defined by the underlying game, institutions are not environmentally determined, but rather are of human construction (but not necessarily of deliberate design).

Because nothing of importance concerning the main points below is lost in taking an especially simple case, I confine myself to the analysis of the evolutionary dynamics governing transitions between two conventions in a two-person two-strategy game in a large population of individuals subdivided into two groups, the members of which are randomly paired to interact in a non-cooperative game with members of the other group. Individuals' best-response play is based on a single-period memory, and they maximize their expected payoffs based on the distribution of the population in the previous period.

The two population subgroups, initially assumed to be of equal size, are termed A's and B's, and each when paired with a member of the other group may choose action 1 or 0, with the A's payoffs, a_{ij} representing the payoff to an A-person playing action i against a B-person playing action

j , and analogously for the B's. If the members of the pair choose the same action they get positive benefits, while if they chose different actions they get nothing. For concreteness, suppose the sub-groups are economic classes selecting a contract to regulate their joint production, which will only take place if they agree on a contract. Payoffs are shares of the joint surplus of the project, with the no-production outcome normalized to zero for both. The payoffs, with the A's as the row player, and the B's as column player, are thus:

To capture the conflict of interest between the two groups, let us assume that $b_{00} > b_{11} = a_{11} > a_{00} > 0$ so the B's strictly prefer the outcome in which both play 0, the A's prefer the equal division outcome which results when both play 1.⁴ Both of these outcomes are strict Nash equilibria, and thus both represent conventions, which I will denote E_0 and E_1 (or $\{0,0\}$ and $\{1,1\}$). Both populations are normalized to unit size, so I refer equivalently to the numbers of players and the fraction of the population, abstracting from integer problems.

Figure 1
Payoffs in the contract game

	B offer contract 1	B offer contract 0
A offer contract 1	a_{11} b_{11}	0 0
A offer contract 0	0 0	a_{00} b_{00}

The state of this population in any time period t is $\{\alpha_t, \beta_t\}$, where α is the fraction of the A's who played 1 in the previous period and β is the fraction of the B's who played 1. For any state of the population, expected payoffs a_i and b_i for the A's and B's respectively playing strategy i , depend on the distribution of play among the opposing group in the previous period, or dropping the time subscript:

$$a_1 = \beta a_{11}; \quad a_0 = (1-\beta)a_{00}; \quad b_1 = \alpha b_{11}; \quad \text{and} \quad b_0 = (1-\alpha)b_{00}.$$

The relationship between the population state and the expected payoffs to each action is illustrated in Figure 2.

Individuals take a given action -- they are 1-players or 0-players -- and they continue doing so from period to period until they update their action, at which point they may switch. Suppose that at the beginning of every period some fraction ω of each sub-population may update their actions (this might be due to the age structure of the population, with updating taking place only at a given period of life, in which case the "periods" in the model may be understood as "generations". Of

⁴ I refer to $\{1,1\}$ as the "equal" convention as a shorthand. The levels of well-being attained by the A's and B's cannot be determined without additional information (if the A's are share croppers who interact with only one B (a landlord), while B's interact with many A's, the "equal" convention would exhibit unequal incomes of the two groups, for example).

course, updating could be much more frequent).⁵ The updating is based on the expected payoffs to the two actions; these expectations are simply the payoffs which would obtain if the previous period's state remained unchanged (the population composition in the previous period being common knowledge in the current period.) While this updating process is not very sophisticated, it may realistically reflect individuals' cognitive capacities and it assures that in equilibrium -- when the population state is stationary -- the beliefs of the actors formed in this naive process are confirmed in practice.

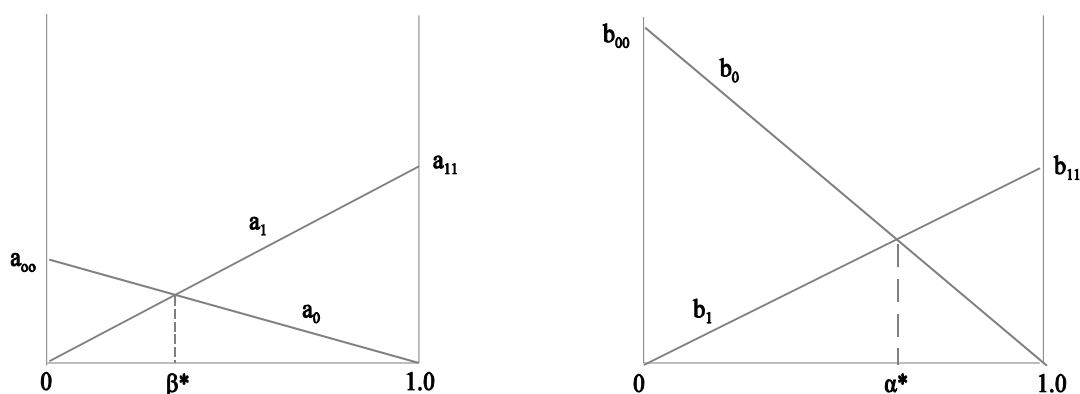


Figure 2 Expected payoffs. Note: A's payoffs depend on β the fraction of B's offering contract 1, while the B's payoffs depend on α the fraction of A's offering contract 1. Because $b_{00} > b_{11} = a_{11} > a_{00}$, the convention E_1 (that is, $\alpha=1=\beta$) is preferred by the A's while E_0 is preferred by the B's.

Individuals are represented simply as bearers of the strategies they have adopted, while the distribution of strategies among them varies. I will analyze the single-period change in the population state ($\Delta\alpha, \Delta\beta$) under the assumption that individual updating of strategies is monotonic in average payoffs so that $\Delta\alpha$ and $\Delta\beta$ have the signs respectively, of $(a_1 - a_0)$ and $(b_1 - b_0)$. The resulting population dynamics are illustrated in Figure 3, where the relevant regions are defined by:

$$(1) \quad \alpha^* = b_{00}/(b_{11}+b_{00})$$

$$\beta^* = a_{00}/(a_{11}+a_{00})$$

⁵ Giving individuals a longer (than one period) memory, or a less naive updating rule, or a more limited knowledge of the distribution of types in the other sub-population, would not yield substantially different insights about the questions explored here. The overlapping-generations assumption concerning updating is, however, important as it means that the stochastic shocks due to idiosyncratic play (to be introduced presently) are persistent as the realized distribution of play in the previous period reflects the shocks experienced over many past periods.

these two population distributions equating the expected payoffs to the two strategies for the two sub-populations, respectively. These values of α and β define best response functions: for $\alpha < \alpha^*$ B's best response is to play 0, and for $\alpha \geq \alpha^*$ B's best response is to play 1, with β^* interpreted analogously.

For states $\alpha < \alpha^*$ and $\beta < \beta^*$ (in the southwest region of Figure 3) it is obvious that $\Delta\alpha$ and $\Delta\beta$ are both negative and the state will move to $\{0,0\}$. Analogous reasoning holds for the northeast region. In the northwest and southeast regions of the state space we may define a locus of states from which the system will transit to the interior equilibrium α^*, β^* , with states below that locus transiting to $\{0,0\}$, and above the locus to $\{1,1\}$. The basin of attraction of $\{0,0\}$, is the area below the dashed downward-sloping line in Figure 3; its size will vary with $\alpha^*\beta^*$. While the interior equilibrium $\{\alpha^*, \beta^*\}$ is an unstable Nash equilibrium (a saddle), the outcomes $\{0,0\}$ and $\{1,1\}$ are absorbing states of the dynamic process, meaning that if the population is ever at either of these states, it will never leave. There being more than one such absorbing state, the dynamic process is *non-ergodic*, that is, its long-run average behavior is dependent on initial conditions.

Chance and Change

How, then, might institutional change occur? Because best response play renders both conventions absorbing states, it is clear that in order to understand institutional change, some kind of non-best-response play must be introduced. Suppose there is a probability ϵ that when individuals are in the process of updating, each may switch their type for idiosyncratic reason. Thus $(1-\epsilon)$ represents the probability that the individual pursues the best response updating process described above. The idiosyncratic play accounting for non-best-responses need not be irrational or odd; it simply represents actions whose reasons are not explicitly modeled in the payoff matrix. Included is experimentation, whim, error, and intentional acts seeking to affect game outcomes but whose motivations are not captured by the above game. Idiosyncratic play can lead to transitions from one convention to another in the following way. If the status quo convention is $\{0,0\}$ but a sufficiently large number of A's play 1 for some reason not captured by the model, then in the next period, the best response of the B's, having encountered these 1-playing A's will be to play 1 as well. In the next period, the best response of the A's who encountered these 1-playing B's will be to play 1, and so on, possibly leading to the "tipping" of the population from the $\{0,0\}$ to the $\{1,1\}$ convention.

For finite populations, the presence of idiosyncratic play transforms the dynamical system described above from a non-ergodic one to an ergodic process with no absorbing states. Ergodicity means that we can specify long term average behavior independently of the initial conditions, a result of central importance in what follows. The simplest case arises when $\omega = 1$ (everyone updates in

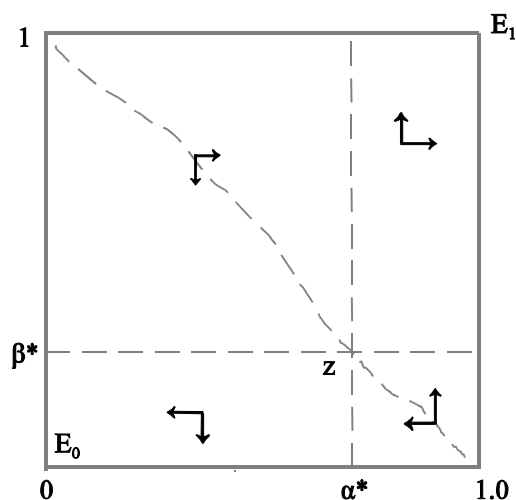


Figure 3 The state space. Note: E_1 and E_0 are absorbing states in the non-stochastic dynamic; z is a saddle.

every period). Then the Markov process described by the model yields a strictly positive transition matrix, meaning that from any state the system will transit to every other state with positive probability. To see that this is true, suppose all members of both sub-populations are "selected" for idiosyncratic play and note that any distribution of their responses is possible, thus giving positive weight to the probability of moving to any state, irrespective of the originating state.⁶ Thus the population state is perpetually in motion, or at least susceptible to movement, and its state is path-dependent: Where it was in the recent past influences where it will most likely be at any moment. History matters, and it never ends.

The fact that the population state is perpetually changing does not mean, of course, that all states are equally likely: the long-run average behavior of the system can be studied. The basic idea is that conventions which require a large amount of idiosyncratic play to dislodge, while requiring little idiosyncratic play to access will persist over long periods, and if eclipsed by some other convention they will readily reemerge. I call these conventions robust. We need to formalize this intuition that robust conventions are "easy to get to, hard to leave."

First, a robust convention is *persistent*: once at or near the convention, it takes a substantial amount of non-best-response play to dislodge it. By *dislodge*, I mean to create a situation in which no further idiosyncratic play is required to lead the population to abandon the convention. Consider the convention E_0 . It can be dislodged in two ways: if more than α^* of the A's or more than β^* of the B's idiosyncratically play 1. The larger are α^* and β^* the less likely is a dislodging event to take place, so these are measures of persistence of E_0 . Likewise, E_1 may be dislodged if more than $(1-\alpha^*)$ of the A's or more than $(1-\beta^*)$ of the B's idiosyncratically play 0.

Second, a robust convention is *accessible*: in the 2x2 case this means the *other* convention is *not* persistent, and it does not require much bunching of non-best-response play at the other convention to displace the population state into the basin of attraction of the robust convention. How accessible is E_0 ? If more than $(1-\alpha^*)$ of the A's or more than $(1-\beta^*)$ of the B's play 0, the population may move from the [1,1] to the [0,0] contract. A bunching of non best response play which tips the population from the basin of attraction of E_1 to the basin of attraction of E_0 is more likely to occur the larger are α^* and β^* , so these are measures of the accessibility of E_0 .

Persistence is analogous to evolutionary stability or non-invadeability introduced by Maynard Smith and Price (1973), α^* and β^* representing *the invasion barrier* or the minimum number of mutant 1-players who would proliferate if introduced into a population of 0-players. Accessibility is analogous to the concept of capacity to invade -- called *initial viability* by Axelrod and Hamilton (1981).)

Note that α^* and β^* thus measure *both* persistence and accessibility of E_0 (with $(1-\alpha^*)$ and

⁶ Where $\omega < 1$ the above intuition remains correct, because if in every period any distribution of play among the potential innovators is possible, then in a sufficiently long period of time any distribution of play among the entire population is also possible.

($1-\beta^*$) the persistence and accessibility of E_1 .) The fact that in the 2x2 coordination game structure the accessibility of a convention is just one minus the persistence of the other will be important below. Thus, if *both* α^* and β^* exceed one half, E_0 has the “easy to get to, hard to leave” qualities of a robust convention. But what if $\alpha^* > 1/2 > \beta^*$, or the reverse? Because there are two ways to get to a convention and two ways to leave, we need to say which of these ways are more likely to occur. I will discuss two answers to this question, one proposed by stochastic evolutionary game theory and the other (to be introduced presently) based on a representation of idiosyncratic play not as accidental but rather as intentional collective action.

Define a *stochastically stable state* as one which occurs with non-negligible probability when the rate of idiosyncratic play is arbitrarily small. As ϵ goes to zero the population will generally spend most of the time at one convention; This is the stochastically stable state. Letting ϵ go to zero solves the problem above of determining which path the population will take in moving from one convention to another: it is more likely to take the most probable path, and as ϵ goes to zero, the probability of taking the less probable path is vanishingly small and hence can be ignored. The more likely path is that which requires fewer cases of non-best-response play.

Following Young (1998), define r_{jk} , the *reduced resistance* on the path from E_j to E_k , as the minimal number of individuals in a population adhering to the convention E_j which, should they idiosyncratically switch their strategy to k , would induce their best-responding partners to switch theirs. Then

$$(2) \quad r_{10} = \min\{(1-\alpha^*), (1-\beta^*)\}$$

$$r_{01} = \min\{\alpha^*, \beta^*\}$$

The convention *to* which the reduced resistance is least is the stochastically stable state. The reduced resistances to a convention are also the risk factors of the convention (r_{jk} is the risk factor of E_k .) So the stochastically stable state is the state with the least risk factor and hence the risk-dominant equilibria.⁷

Thus the convention $\{0,0\}$ will be stochastically stable if

$$r_{10} = \min\{(1-\alpha^*), (1-\beta^*)\} < \min\{\alpha^*, \beta^*\} = r_{01}$$

Using the payoffs $b_{00} > b_{11} = a_{11} > a_{00}$ we have

$$(3) \quad r_{10} = (1-\alpha^*) = 1 - \{b_{00}/(b_{11}+b_{00})\} = \{b_{11}/(b_{11}+b_{00})\}$$

⁷ Young (1998), theorem 4.1. In the updating model on which this theorem is based (and the Contract Theorem below) agents have a memory of m periods, and sample ($s < m$) from their memory to form expectations. (In the model in the text $s=m=1$.) Young's results concerning stochastic stability generalize beyond the 2x2 coordination games treated here.

$$r_{01} = \beta^* = a_{00}/(a_{11}+a_{00})$$

Thus as ε goes to zero it is the idiosyncratic actions of the B's that propel a movement from [0,0] to [1,1] while the A's idiosyncratic actions induce the reverse tipping. The convention [0,0] will be the stochastically stable state if $(1-\alpha^*) < \beta^*$, or using the above expressions,

$$(4) \quad a_{00}b_{00} > a_{11}b_{11}$$

Note that the two terms in (4) are just the product of the difference between A's and B's payoffs and their fallback position (which is zero).. Thus a contract that is closer (in this sense) to the Nash solution for the division game is the stochastically stable state. This should come as no surprise given that the Nash solution is the stationary distributional norm in a plausible dynamic with occasional idiosyncratic play.

What does (4) tell us about the characteristics of stochastically stable states? Suppose contracts differ in their distributional shares and also in the level of total surplus (sum of payoffs) they yield. Some contracts are, in this sense, more efficient than others. This might occur if the use of a particular technology required a distinct set of property rights, which in turn supported a particular equilibrium contract. An example of this technology-contracts mapping was seen in the case of the rise of agriculture and the emergence of individual property rights. Analysis of the 2x2 contract game will be facilitated if we write $a_{11}=1$, $b_{11}=1$ and $a_{00} + b_{00} = \rho$, so $\rho/2$ is a measure of the relative efficiency of the {0,0} convention; when ρ takes the value of 2, the two conventions produce the same the joint surplus. Further let the A player's share of joint surplus in the B-favoring {0,0} equilibria be $\sigma < 1/2$, with $(1-\sigma)$ the share of gained by B.

To explore the effect of the terms of the contract on the stochastic stability of the state defined by the convention in which that contract is universal, consider the contract space in Figure 5. The {1,1} contract is defined as the Benchmark Contract, with E_1 the associated convention. The contract space depicts a set of Alternative contracts defining convention E_0 . Point S' is the Benchmark contract (with $\rho = 2$ and $\sigma=1/2$). Thus if the two possible contracts are represented by points S' and x , both groups will prefer the Alternative contract because both $\sigma\rho$ and $(1-\sigma)\rho$ exceed 1 under its terms. Contracts above AS' are Pareto-superior to the benchmark. (Ignore the locus SS' for the moment.)

Figure 4
Modified payoffs in the contract game

	B offer Contract 1	B offer Contract 0
A offer Contract 1	$a_{11}=1,$ $b_{11}=1$	0
A offer Contract 0	0	$a_{00}=\sigma\rho,$ $b_{00}=(1-\sigma)\rho$

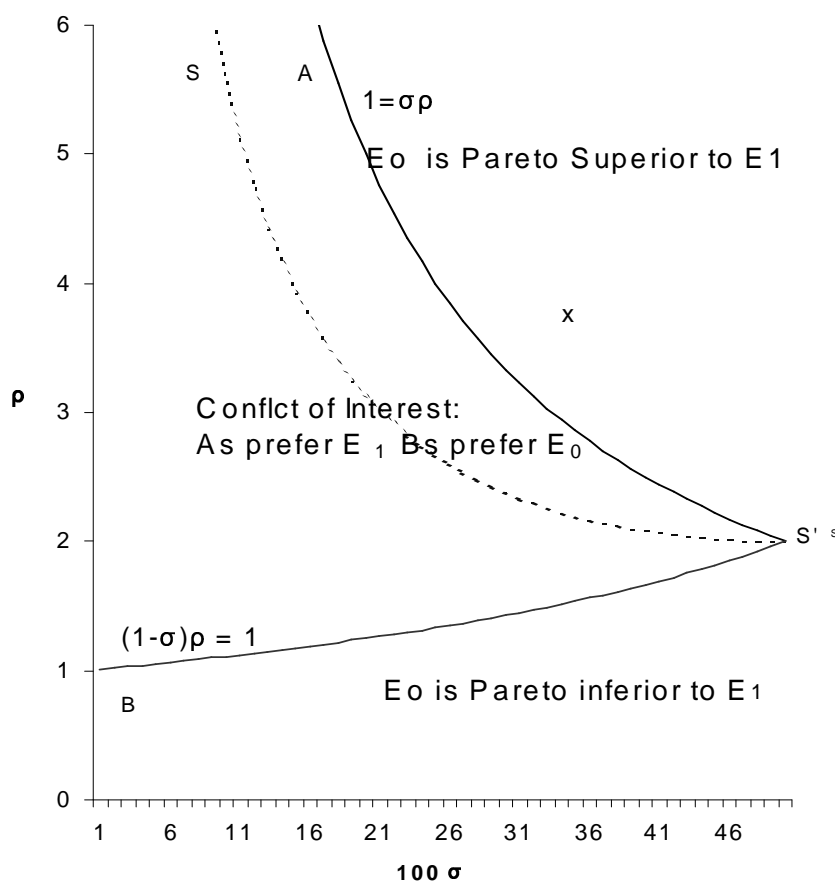


Figure 5 Contrasting contracts. Each point represents the efficiency and distributional share of the Alternative contract supporting the equilibrium E_0 . Contracts above $S'A$ are Pareto superior to the Benchmark contract with $\rho = 2$ and $\sigma = \frac{1}{2}$. Contracts below $S'B$ are Pareto-inferior to the Benchmark contract.

Conflict of interest between the two groups is confined to the contracts lying below AS and above BS . This does not ensure that the S' would be eclipsed by an alternative contract like x . The reason is that while x is Pareto superior to the S' , adherence to S' is a mutual best response and so will only be dislodged by non-best-response play. Our intuition, however, says that Pareto inferior conventions must be at a disadvantage in a stochastic environment. Our intuition is correct: Pareto inefficient conventions are not robust in this evolutionary dynamic, and we can say considerably more.

A striking theorem due to Peyton Young (1998) demonstrates that the institutions supporting stochastically stable states are not only efficient but also egalitarian, if we give this term a rather special meaning. For any two contracts call the *relative payoff*, π_{ij} the payoff to members of group i in contract j , relative to the maximum payoff they get in either of the two contracts. Under some

innocuous restrictions on the updating process, Young's "Contract Theorem" shows that the stochastically stable state is the one which maximizes the relative payoffs of the group with the lowest relative payoff.⁸ Why this is true, and the sense in which the property that stochastically stable states are maximin in relative payoffs can be termed egalitarian will be clarified by making use of what we already know about these states.

The convention $\{0,0\}$ will, as we have seen, be the stochastically stable state if $a_{00}b_{00} > a_{11}b_{11}$, which using the payoffs in Figure 4 requires that

$$(5) \quad \sigma(1-\sigma)\rho^2 > 1$$

It is clear from this condition that both relative efficiency and equality of shares contribute to stochastic stability of a convention (the term $\sigma(1-\sigma)$ is maximized for $\sigma = 1/2$.) Figure 5 illustrates the relationship between efficiency and equality as determinants of stochastic stability: SS' is the locus of combinations of ρ and σ such that $\sigma(1-\sigma)\rho^2 = 1$ and which thus equate the risk factor of $\{0,0\}$ to the risk factor of the egalitarian convention $\{1,1\}$ (for which $\rho=2$ and $\sigma=1/2$). Thus SS' is the locus of Alternative contracts such that both conventions are stochastically stable. Alternative contracts above SS' are stochastically stable when the other convention is based on the Benchmark contract. For Alternative contracts below SS' the Benchmark contract is stochastically stable.

Note that while stochastically stable states are maximin in *relative* payoffs, they are not maximin in payoffs. Alternative contracts lying between $S'S$ and $S'A$ are stochastically stable, but the payoffs of the A's are lower in the Alternative contract than in the Benchmark contract. Thus stochastically stable states are egalitarian only in a rather special sense.

It is easy to see why efficient conventions would be favored in this setup. For at least one group, offering the efficient contract must be risk dominant in the standard sense that if one believes that the other will offer the two contracts with equal probability, then the best response is to offer the more efficient one. Inefficient conventions are not accessible because it takes a large amount of non-best-response play to induce best responders to shift from an efficient to an inefficient convention. Note that this is not because best responders anticipate the consequences of their switching for the population level dynamics. Rather, their response is purely individual and based on past (not anticipated future) population states; no individual is attempting to implement the more efficient convention. Inefficient conventions are not persistent for analogous reasons.

⁸ To see that stochastically stable states are maximin in relative payoffs it is sufficient to show that the condition $\sigma(1-\sigma)\rho^2 = 1$, which defines equivalent stochastic stability of the Alternative and Benchmark contracts also equates the minimum relative payoffs of the two contracts. Consider an Alternative contract such that both contracts are stochastically stable. Then we have $\pi_{A0} = \sigma\rho < 1 = \pi_{B0}$ and $\pi_{B1} = \rho(1-\sigma)^{-1} < 1 = \pi_{A1}$ and the minimum relative payoff in the Alternative and Benchmark contracts respectively are $\sigma\rho$ and $\rho(1-\sigma)^{-1}$. Equating these gives the above condition for the states given by the two contracts both being stochastically stable.

Less transparent is the result that highly unequal conventions are not good candidates for stochastic stability. This is a consequence of the fact that they are easily unraveled, because as Young (1998):137 puts it: "it does not take many stochastic shocks to create an environment in which members of the dissatisfied group prefer to try something different." Note that in this example, as in the discussion of reduced resistances above, it is the idiosyncratic play of the *privileged* group that unravels the unequal convention, that is, the convention from which they benefit disproportionately.

To see why the processes of transition between the two conventions depends on the share of the less well off in the unequal convention, we can use (3) and the data in Figure 4 to get the following expressions for the reduced resistances on the paths to the two equilibria.

$$r_{10} = (1/\{1+(1-\sigma)\rho\})$$

$$r_{01} = \sigma\rho/(1+\sigma\rho)$$

As σ goes to zero (the poor get nothing in the unequal convention), the resistance on the path to the equal convention (r_{01}) also goes to zero. The reason is that in a population near the $\{0,0\}$ convention, even if the A's (the poor) believed that virtually all of the B's would play 0, their best response would nonetheless be to play 1. The reason is that for $\sigma = 0$, they would not benefit from concluding a contract with a 0-playing B, so as long as there was some chance of meeting a 1-playing B expected payoffs would be maximized by playing 1. Thus the population will transit to the more equal convention for an arbitrarily small amount of non best response play by the rich. This is the evolutionary game theorist's rendition of Marx's rhetoric about the working class having "nothing to lose but its chains." Thus the unequal convention becomes less persistent as it becomes more unequal.

Figure 6 shows that more unequal shares in the $\{0,0\}$ convention makes both conventions more accessible(that is, it reduces the resistance to both equilibrium). But the accessibility of the more equal convention is increased relatively more. The reason why $\{0,0\}$ becomes more accessible is that in the neighborhood of the $\{1,1\}$ convention, it takes fewer non-best responding A's to induce the B's to take a chance and play 0 (if they happen to meet a 0-playing A, they will do very well.). Thus, the resistance on the path to the unequal convention falls as σ falls. But resistance on this path remains positive even when the B's get all of the joint surplus in $\{0,0\}$ for in this case $r_{10} = 1/(1+\rho)$.

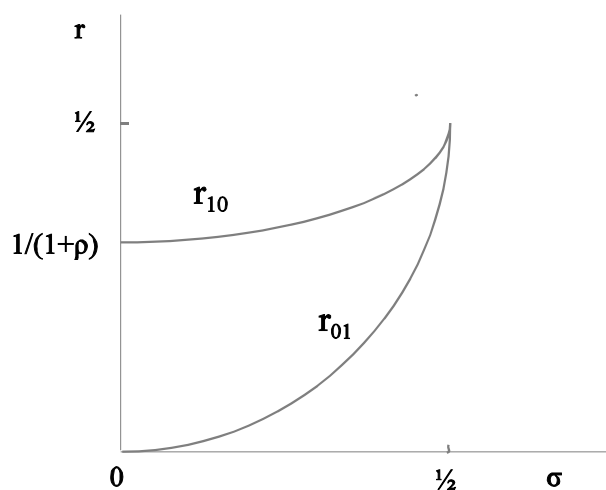


Figure 6 Risk factors and the degree of inequality. Note: greater inequality of shares in the unequal convention (lower σ) reduce the risk factors of both conventions, but affect the equal convention more than the unequal one.

I have illustrated the insights of stochastic evolutionary game theory using a comparison of just two contracts; but note that *any* two contracts along the SS' locus in Figure 5 are both stochastically stable states. We may thus interpret SS' as an “iso-stochastic stability” locus, and note that this is just one of a family of such loci. For any two contracts i and j along one of these loci it is the case that $a_{ii} b_{ii} = a_{jj} b_{jj}$. Now suppose, that given the technologies, preferences and other relevant data obtaining in some historical period, there is a set of feasible contracts defined in $[\rho, \sigma]$ space. Two members of the family of iso-stochastic stability loci ($S'S'$ and S'',S'') and the feasible contract set bounded by CC are illustrated in Figure 7. If only two contracts are considered, x , and y , we would expect the population to move between these two conventions in the very long run, spending equal amounts of time at each. But if x were the current convention and z the alternative, then we would expect z to emerge and to persist virtually all of the time.

The advance of technology and the evolution of preferences takes the form of a shift in the feasible contract set. One possible such shift induced by the introduction of a new technology for which a larger share for the B's and a smaller share for the A's is appropriate is indicated by the new contract possibility frontier $C'C'$. Stochastic evolutionary game theory would lead us to expect a new contract to emerge, one with a reduced sigma, indicated by z' at the tangency of the new contract possibility frontier and a higher iso stochastic stability locus. A process of this type may have occurred with the introduction of agriculture, or the development of capitalism half a millennium ago.

The introduction of idiosyncratic play removes the deterministic dependence of outcomes on initial conditions which characterizes the non-stochastic approach. Rather, the stochastic approach allows predictions of the average population state over a sufficiently long historical period, along with a rather strong characterization of the nature of these stochastically stable states. The approach thus provides one account of how the institutions which Parsons termed “evolutionary universals” might come to be recurrent historically and ubiquitous at any given point in time: Institutions supporting stochastically stable states would have been, as Parsons (1964):340 put it, “likely to be ‘hit upon’ by various systems operating under different conditions” and to persist over long periods.

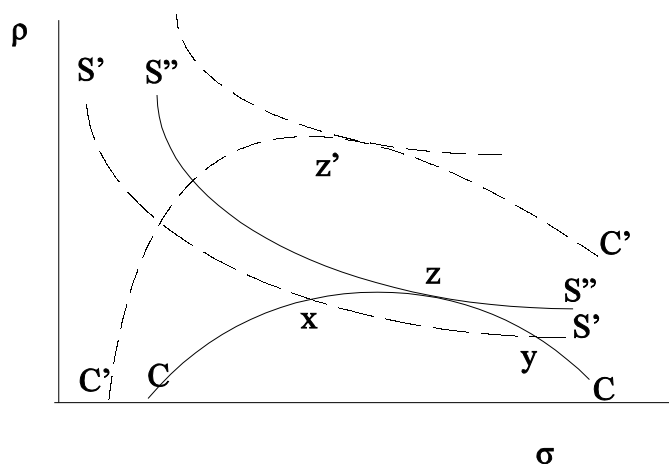


Figure 7 Equilibrium selection by chance from a feasible contract set

Intentional Non-best-response Actions with Sub-populations of Different Size

Stochastic evolutionary game theory makes two major contributions to the study of institutional dynamics. First, it allows us to go beyond the correct but not very illuminating

conclusion that “history matters” and to study how evolutionary processes will favor some kinds of institutions over others. Second, it provides a way of taking account of the importance of chance events. The major shortcomings of stochastic evolutionary game theory as an account of real historical processes of institutional change are two. First, the relevant theorems about the characteristics of robust institutions apply only when the rate of non-best response play is arbitrarily small.⁹ Second, it neglects the important part played by collective action in the process of institutional innovation and transformation. It was not a fortuitous piling up of unlikely accidents that doomed apartheid or Communism, but rather a combination of chance events and the deliberate and coordinated actions taken by those seeking to live under other institutions.

When intentional non-best-response play is introduced in the form of collective action by those trying to displace the status quo convention, the dynamic of institutional innovation is substantially altered. It is no longer generally the case that stochastically stable states are egalitarian and efficient. In particular, if the rich are few and the poor many unequal and inefficient institutions can be very robust. The reason is that when non-best-response play is both intentional and non-negligible, there is just one way (rather than two) that a convention can be overturned (by the actions of those who would benefit more at the other convention) and the larger numbers of the poor militate against a sufficient fraction of them adopting a non best response to displace the equilibrium under which they do poorly.

The collective action approach requires some modifications in the above model. First the players must be assumed to recognize the possibility of transiting to a new institutional setup, and have the ability to anticipate the consequences of their actions on the actions of others. Thus rather than restricting individuals to backward-looking updating, I now introduce a limited capacity to look forward. Second, when the frequency of idiosyncratic play is non-negligible, the reduced resistances introduced above no longer provide the basis of an account of institutional transformation. The reason is that their relevance is based on non-best response play being sufficiently infrequent that the least probable of the two paths from one convention to another can be ignored. Rather than letting ϵ go to zero, the approach below identifies probable paths from one convention to another by endogenizing the process of idiosyncratic play using of a model of collective action.

By *collective action*, I mean the intentional joint action towards common ends by members of a large group of people who do not have the capacity to commit to binding agreements prior to acting (that is, they act non-cooperatively). Examples are strikes, ethnic violence, insurrections,

⁹ Young (1998) shows that for a single population 2x2 game the population spends most of the time at the stochastically stable state even when ϵ is substantial (e.g. 0.05, or even 0.10) as long as the population is large (and hence transitions infrequent even with substantial non best response play.) Note that in this single population 2x2 case there is just one way to transit from one convention to the other, so this result is not very surprising. By contrast, in the two-population game, letting ϵ go to zero selects which of the two paths from one convention to the other is to be the basis of the calculation. With substantial error rates both paths must be considered (because the least probable path may be followed with substantial likelihood).

demonstrations, and boycotts. An individual's participation in a collective action may be modeled as an idiosyncratic non-best-response, one which does not take the form of stochastically generated "errors" but instead represents an intentional action motivated by the desire to improve one's well-being and perhaps the well-being of others. For this reason it is likely that the extent of non-best response play will co-vary among individuals and depend on the payoff structure and other aspects of the pattern of social interaction defining the underlying game.¹⁰

To clarify the underlying processes, I will first analyze a degenerate case in which individuals participate in a non-best-response collective action when it is in their individual interest that the action take place. Suppose that everyone updates in each period ($\omega = 1$) and assume that there is a probability $\varepsilon \in (0,1)$ that each person is "called to a meeting" at which those attending consider undertaking a non-best-response action. For example, assume the B-favorable convention $\{0,0\}$ obtains and some fraction of B's (resulting from the "call") are considering switching to offer a 1-contract instead. But they cannot benefit from switching because they prefer the status quo convention, and destabilizing it -- should sufficiently many of the other possible B-innovators also switch -- could propel them to the alternate convention under which they would be worse off. These potentially idiosyncratic players would thus decline the opportunity to innovate.¹¹

By contrast, imagine that the a group of A's were randomly called for deliberation of the merits of a switch away from the governing convention $\{0,0\}$, and suppose that should they all adopt a non-best-response, this will be common knowledge. Each then might reason as follows. If they are sufficiently numerous and if all of them switched, the best response for the B's would be to switch as well. Knowing this, should they all switch, they would anticipate the B's response and so would persist in offering 1-contracts in the next period. As a result, the A-unfavorable convention $\{0,0\}$ would be displaced.

Suppose there are n members of the A population (previously normalized to unity). Because if fewer than $n\alpha^*$ A's were called, there could be no benefit to collective action even if it were uniformly successful, let us analyze the case for which the number called, η , exceeds this critical level, that is $\eta \geq n\alpha^*$. To lend some concreteness to the case let us say that switching means to engage

¹⁰ Bergin and Lipman (1996), Young (1998) and {van Damme, 2001 #3588} analyze state dependent mutations. The proviso that play is non-cooperative excludes the degenerate case (with which I begin for purposes of illustration) of groups whose structure allows the assignment of obligatory actions to each of its members. While most successful collective actions include a wide range of selective incentives and sanctions to deter free riding, few if any groups have the capacity to simply mandate group-beneficial behaviors by individual members.

¹¹ Favored groups, like the B's in convention $\{0,0\}$ may deploy informal or governmental sanctions or to minimize idiosyncratic play of their own members. Examples include the shunning and more severe sanctions imposed on whites offering favorable contracts to non whites in racially stratified societies such as apartheid South Africa and the U.S. South prior to the civil rights movement.

with other A's in a strike, refusing to accept any outcome less than a_{11} (all this means is to offer a 1-contract, so the strategy set is unchanged). We can explore the long run behavior of the system by calculating τ_0 , the expected waiting time (number of periods) before a strike by the A's induces a transition from convention $\{0,0\}$ to $\{1,1\}$. This is the inverse of the probability, μ_0 , that in any period a transition from $\{0,0\}$ will be induced or $\tau_0 = 1/\mu_0$. To determine this probability one may proceed as follows. First, count the subsets of A's sufficiently numerous to induce a transition, then determine the probability (given ϵ) that each subset will be drawn; then sum these probabilities to get the probability that any transition inducing event occurs, μ_0 . In this degenerate case of assured collective action when it is beneficial, any subset of A's with $n\alpha^*$ or more members will induce a transition. So using $C_{n,n\alpha^*+i}$ to indicate the number of subsets of m members in a population of n individuals we have

$$\mu_0 = \sum C_{n,n\alpha^*+i} \epsilon^{n\alpha^*+i} (1-\epsilon)^{n-n\alpha^*+i} \quad \text{for } i = 0 \dots n(1-\alpha^*)$$

An example will clarify the calculation. Suppose $\epsilon = 0.1$, four individuals (W,X,Y, and Z) make up the A sub population, $\alpha^* = 3/4$. Then the A-unfavorable convention E_0 will be displaced by idiosyncratic play by any of the following combinations: WXY, XYZ, YZW and WXYZ. The first three of will each occur with probability 0.0009 and the last with probability .0001, so $\mu_0 = .0028$ and $\tau_0 = 357$ periods. As we want to know the long run average behavior of the system, we calculate τ_1 in a manner analogous to the above and express the average time at or near E_0 , λ_0 as

$$\lambda_0 \equiv \tau_0 / (\tau_0 + \tau_1)$$

with $\lambda_1 \equiv 1 - \lambda_0$. If there are three B's and $1 - \beta^*$ (the critical fraction required to displace the B-unfavorable convention E_1) is $2/3$ then $\mu_1 = .028$ and $\tau_1 = 35.7$ periods, so $\lambda_0 = 0.90$

Figure 8 gives the results of this calculation where the two sub-populations each have 12 members and for various values of ρ and σ . Where E_0 is identical to E_1 ($\rho = 2$ and $\sigma = 1/2$ indicated by the dark bar at these coordinates) the population spends half of its time at each convention. One can see a band of conventions (similar to the locus SS' in Figure 5) which like ($\rho = 2$ and $\sigma = 1/2$) generate equal average waiting times (for example, $\rho = 2.5$ and $\sigma = 0.2$ generates this result, as does $\rho = 2.25$ and $\sigma = 0.3$). The population will spend virtually all of the time at conventions more efficient or more equal than these and virtually none of the time at conventions less efficient or less equal.

The reason that more equal conventions are favored in this framework is the following. Consider an Alternative contract with $\rho = 2$ and $\sigma < 1/2$. An increase in the distributional share of the A's in the Alternative contract has two effects. First, it lowers α^* and thus it requires fewer instances of idiosyncratic play by the A's to disrupt the Alternative Contract, inducing a movement to the Benchmark (which they prefer). The reason is that when the Alternative is less unequal, it takes fewer idiosyncratic A's to induce the B's to switch to the Benchmark. The second effect of an increase in σ is to raise β^* thus reducing the minimal fraction of non best responding B's (namely $(1 - \beta^*)$) required to induce the A's to abandon their preferred Benchmark contract in favor of the Alternative. The two effects of a more equal Alternative contract work in opposite directions, the

first leading to a shorter waiting time for a transition from the Benchmark to the Alternative, and the second leading to a shorter waiting time for the reverse transition. But for $\sigma < 1/2$ the second effect is larger, so the population will spend more time at the Alternative, the more equal it is.

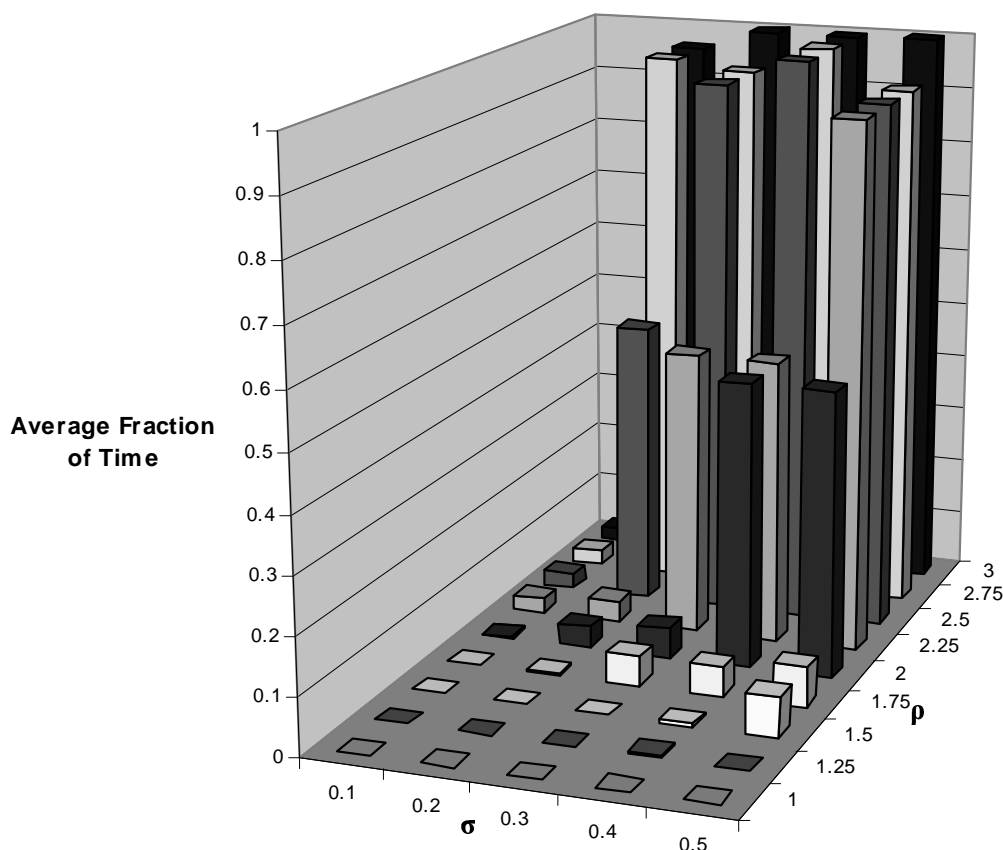


Figure 8 Efficient and equal conventions are stochastically stable with equal sub-population sizes. Note: the Benchmark convention is E_1 for which $\rho = 2$ and $\sigma = 1/2$.

Note that Figure 8 confirms that despite the restriction of non-best-response play to group-beneficial actions, the system will spend most of its time in the stochastically stable states. This may seem remarkable given that the transitions governing the dynamic in the stochastic evolutionary approach are that the B's idiosyncratic play disrupts the B-favorable convention and similarly for the A's. By contrast, the collective action approach dismisses these transitions as irrelevant, focusing instead on non best response play motivated by the prospect of increasing one's payoffs by inducing an institutional transition, idiosyncratic play by the A's disrupting the B-favorable convention, and conversely.

Why is the long-run average behavior of the system is not affected by substitution of intentional collective action for the elimination of the least probable path as ϵ goes to zero? The reason is that convention E_0 is more vulnerable to intentional collective action (by the A's) than E_1 (by the B's) if $\alpha^* < (1-\beta^*)$ while abstracting from intentions (that is permitting the idiosyncratic play

of those benefitting from a convention to displace it), E_1 is the stochastically stable state if $\beta^* < (1 - \alpha^*)$ and the two conditions are equivalent. Thus, the same state is identified as the more robust by the two measures. But this is a special result of the 2x2 game structure and it does not generalize to larger games, or as we will see, to 2x2 games with a more realistic (non-degenerate) process of collective action, and to cases in which the two sub-populations are of different size.

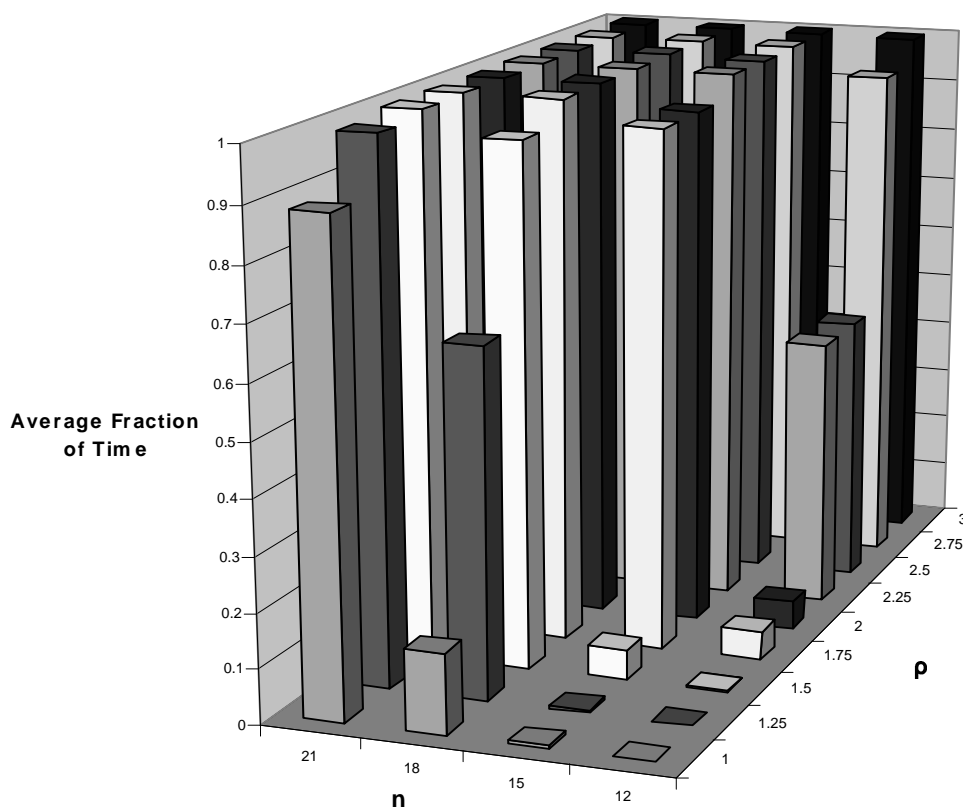


Figure 9 Unequal conventions persist when the poor outnumber the rich.

Note: total population is 24; the Benchmark convention is E_1 ($\sigma = 1/2$, $\rho = 2$). E_0 is characterized by the values of ρ indicated and $\sigma = 0.3$.

Figure 9 shows the effect of assuming sub-populations of different size (retaining the degenerate model of collective action) for an alternative contract with $\sigma = 0.3$ and with the ρ values as shown. By contrast to the equal sub-population size case depicted in Figure 8, when population sizes differ the intentional nature of non-best-response behavior makes a difference: unequal and quite inefficient conventions may be highly persistent. For example, in the equal population size case a convention with $\sigma = 0.3$ needed a ρ of 2.25 to be equally persistent to E_1 ; but if the A's number 18 and the B's 6, the two conventions are equally persistent when the unequal convention is much *less* efficient than the benchmark, that is $\rho = 1.25$. Where there are 21 A's (and 3 B's) the population will spend most of the time in the unequal convention even if its level of efficiency is half that of the equal convention. Note that the level of inequality measured by the average income of B's relative to A's is $n(1-\sigma)/\sigma(24-n)$, each B interacting with more A's as their relative share of the population increases. Thus at the convention E_0 if $\sigma = 0.3$ and the A's and B's are equally numerous, the B's have

an income 2.33 times the A's but when there are 21 A's and 3 B's, the ratio is 16.33. Thus highly unequal distribution of income may result from unequal sub-population sizes, and may be persistent because of the unequal sub-population sizes.

The evolutionary success of unequal and inefficient conventions benefitting the smaller of the two classes is readily explained. As long as rate of idiosyncratic play is less than the critical fraction of the population required to induce a transition (which I assume), smaller groups will more frequently experience “tipping opportunities” when the realized fraction of the population who are “called” by chance exceeds the expected fraction (ϵ itself). The theory of sampling error tells us that the class whose numbers are smaller will generate more “tipping” possibilities. Small size does not facilitate collective action if more than the critical number are “called”: recall that in this case, all of those called will choose the risk dominant strategy, and this is independent of their numbers.

Collective Action

So far I have abstracted from the problem of collective action by assuming that whenever a sufficient fraction of a sub-population is “called” they will adopt a non-best-response if they (and their group) would benefit if all of those called adopted the non-best-response. Extending stochastic evolutionary game theory this means imposing a particular social structure on the process generating non-best-response play. This structure must explain why actions which are non-best-responses in the contract game may nonetheless be the result of intentional action when the game is amended to include the possibility of collective action. Thus, what is needed is a model of the coordination problem posed by collective action, nested in the larger population game representing institutional evolution. Taking account of both the intentional nature of collective action and the coordination problem peculiar to it will augment the stochastic approach in illuminating ways.

Because collective actions generically take the form of n -person public goods games in which the dominant strategy is non-participation if preferences are wholly self-regarding, the extended model must address incentives for each to free ride when other act in pursuit of commonly shared objectives. A second desideratum is that the model should reflect the fact that opportunities for collective action often arise by chance, or at least in ways too complex to tractably model, examples being economic depressions, wars, price shocks, booms, and natural disasters. Finally, unlike idiosyncratic play, participation in collective action is not only intentional (rather than accidental) but is also conditional on one's beliefs about the likelihood and consequences of a substantial number of one's kind changing behaviors. For this reason, facts about global rather than simply local payoff (that is, payoffs both in the present convention and in the alternative, rather than those in the neighborhood of the current population state alone) may have a bearing on the outcomes.¹²

Engaging in this collective activity yields in-process benefits of two types. First, irrespective of the consequences of the action, conformism (or punishment of non-conformists) may impose a

¹² This means that individuals are forward looking to the extent that they can anticipate the consequences of successful collective action.

cost on those not adopting the most common action. So, let c be the cost of being a sole non-conformist, and the conformism costs to those striking being $(1-s)c$ where s is the fraction of those “called” who strike. The costs to the non strikers is sc . Further, there are net benefits or costs associated with the action that may be independent of the numbers participating, including both the time, resources and possibly risk of harm associated with the collective action as well as the positive value of participating, or what Wood (2003) terms the "pleasure of agency".¹³

It is reasonable to suppose that these subjective benefits depend on the magnitude of the gains to be had if the action is successful, not primarily because these gains are a likely consequence of one's individual participation (which is very unlikely in large groups) but because the magnitude of the gains to be had is plausibly related to the strength of the norms motivating the action. The pleasure of participating in a collective action that would if successful transform the conditions of one's class from squalor to abundance is likely to be greater than the pleasure of acting for wage increase a few cents more an hour.

So let the net subjective benefits for an A engaging in a collective action to displace convention $\{0,0\}$ be

$$(6) \quad \delta = \delta(a_{11} - a_{00})$$

where δ is a positive constant, reflecting the fact that joining a collective action in pursuit of an institutional change from which one and one's peers will not benefit confers no benefits.¹⁴ If the strike fails (because too few participate in it) the status quo convention will persist, and all A's will get a_{00} in subsequent periods independently of whether they participated in the strike or not. Likewise if the strike succeeds all As will get a_{11} subsequent periods, irrespective of their actions this period. Thus the relevant comparison is between the single period net benefits to striking (insisting on contract 1, refusing contract 0) or abstaining are:

$$(7) \quad u_1 = \delta(a_{11} - a_{00}) - (1-s)c$$

$$(8) \quad u_0 = a_{00} - sc$$

These payoff functions are illustrated in Figure 10, from which it is clear if those involved believe

¹³ Compelling evidence from the histories of collective action (e.g. Moore (1978)) anthropology (Boehm (1993), Knauff (1991)) and experimental economics suggests that individuals knowingly engage in costly actions to punish violations of norms, even when these actions cannot otherwise benefit the individual.

¹⁴ Conventions typically not only allocate gains but also influence the cultural and political conditions relevant to the net costs and benefits of engaging in collective action. But I here abstract from this (the δ 's are not subscripted to indicate the convention defining the status quo ante.)

that at least s^* of their fellows will join in, then strikers' expected payoffs will exceed those of non-participants, and hence all will elect to strike. The critical value, s^* equates u_0 and u_1 :

$$(9) \quad s^* = \frac{1}{2} - [\delta(a_{11} - a_{00}) - a_{00}]/2c$$

How might A's beliefs be formed? The simplest supposition consistent with the above model is that having no information about what the others will do, each believes that the likelihood of each of the others participating is $\frac{1}{2}$, so the expected fraction participating is $\frac{1}{2}$, and all will participate if s^* less than one-half.¹⁵

Thus unanimous participation (of those “called”) will occur if striking is the risk dominant equilibrium of the collective action game, requiring that the numerator of the bracketed term on the right hand side of (9) be positive, or that the “pleasure of agency” outweighs the loss of a single period’s income. (Note that while inferior payoffs in the status quo convention ($a_{11}-a_{00} > 0$) is a necessary condition for participation, it is not sufficient, as it does not insure that $\delta(a_{11}-a_{00})-a_{00} > 0$.)

The properties of the dynamical system are substantially altered by modeling idiosyncratic play as intentional collective action. Notice that if $\delta(a_{11}-a_{00})-a_{00} < 0$ collective action will not take place (irrespective of the numbers of randomly drawn potential innovators), so the A-unfavorable convention $\{0,0\}$ is an absorbing state. Thus the dynamical system with collective action as the form of non-best-response play is non-ergodic, and institutional lock-ins are possible, with initial conditions determining which of the two conventions will emerge, and then persist forever. To see that this must be the case for a finite “pleasure of agency” parameter δ , consider an unequal convention with $a_{11}-a_{00} = \Delta$ letting Δ become arbitrarily small must make $\delta(a_{11}-a_{00})-a_{00} < 0$ so collective action by A’s will not occur and E_0 , should it ever occur, will persist forever. Thus there must exist a set of conventions, less equal than E_1 and no more efficient, that are absorbing states.

¹⁵ The choice of $\frac{1}{2}$ is conventional but arbitrary; individuals may have prior beliefs of the fraction likely to participate based on previous similar situation and the like. If individuals then apply their reasoning to each of the others (each, supposing that half will participate, will also participate), they would then correctly predict that $s=1$; but while this second round of induction may determine whether the individual expects the collective action to be successful in displacing the convention, is not relevant to the individual's behavior, as the relative payoffs of participating or not are independent of the success of the action.

The collective action model thus gives a quite different dynamic than the conventional stochastic evolutionary model. Figure 11 reproduces the contract space for the Alternative Contract for the case where $\delta=2$ (the Benchmark, E_1 being $[1,1]$ and SS' the locus of Alternative contracts that are equally stochastically stable to the Benchmark). Very efficient or very equal Alternative contracts are absorbing for they are either Pareto-superior to the Benchmark or at least provide sufficient benefits to the A's to preclude their taking collective action. It can be seen that E_0 may be absorbing even if it would not have been stochastically stable in the conventional model. For the region where neither contract is absorbing, the long term average behavior summarized in Figure 9 applies; here the size of the two sub-populations matters.

How are we to interpret the absorbing states? Over relevant time scales, the parameters of the model are likely to change due to cultural and political changes affecting δ or technical or other changes affecting the payoffs to the relevant contracts. Suppose some unequal Alternative contract defines the status quo convention (E_0), and it represents an absorbing state. If technical change made the $\{1,1\}$ contract progressively more efficient by comparison to $\{0,0\}$, then $\delta(a_{11}-a_{00})$ would eventually exceed a_{00} . As a result, the conditions for collective action would obtain, and a transition from E_0 to E_1 would eventually take place. Transitions in the reverse direction would become more unlikely over time as the increase in a_{11} raises the minimum number of non-best-responding B's required to unravel E_1 . Thus, the institutional demands of new technologies may account for the emergence of new contractual conventions. A cultural change enhancing the pleasure of agency, δ – a role played by liberation theology in some parts of Latin America, by the spread of democratic ideology in South Africa and the former Communist countries - would have the same effect.

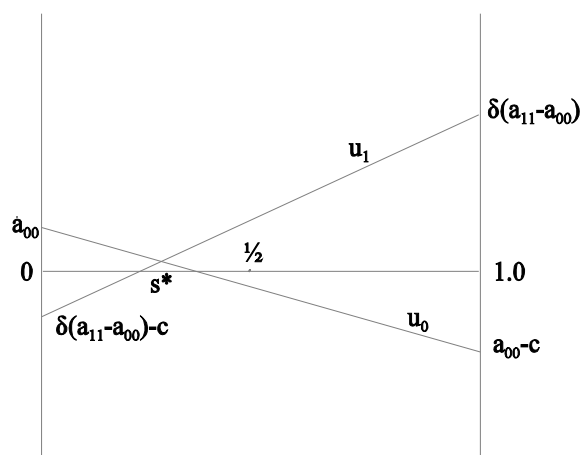


Figure 10 The collective action problem.
 Note if $s^* < 1/2$ the risk dominant equilibrium is universal participation in the non-best-response action (from Wood, 2003)

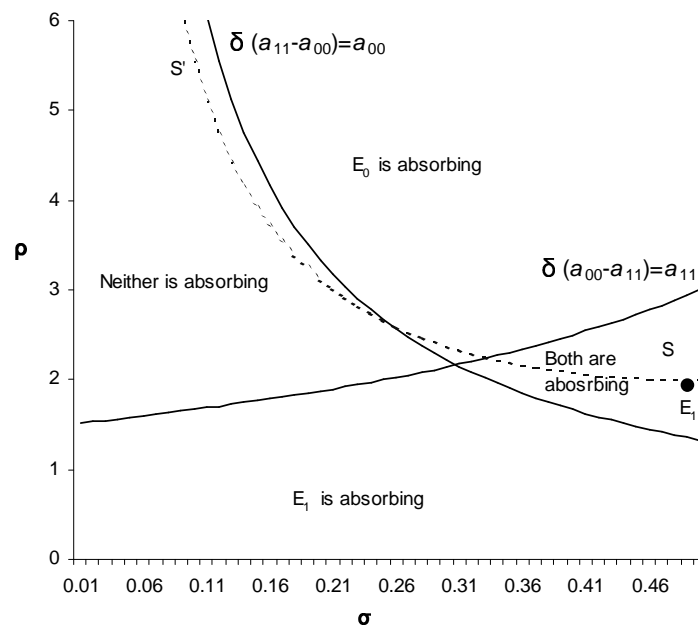


Figure 11 Equilibrium selection by chance and collective action. Note: the long term behavior of the population in the region for which neither convention is absorbing is illustrated in Figure 9.

This is very roughly Marx’s account, which sees history as a progressive succession of “modes of production,” each contributing to “the development of the forces of production” for a period, then becoming a “fetter” on further technological advance and being replaced through the collective action of the class which would benefit by a shift to a new convention more consistent with the new technologies.

Conclusion: the institutional ecology of inequality

The integration of chance and collective action developed here is far from the first proposed marriage of Darwin and Marx. Writing to Engels in 1860, Marx saw parallels between *The Origin of Species* and their *Volume I of Capital*, both of which had been published the previous year: “Although it is developed in the crude English style, this is the book which contains the basis in natural history for our viewpoint” (Padover (1979):139). Fourteen years later at Marx’s grave side, Engels would say: “Just as Darwin discovered the law of evolution in organic nature, so Marx discovered the law of evolution in human society.”

Stochastic evolutionary game theory has recently made available powerful analytical tools of Darwinian inspiration, providing an illuminating framework in which to consider the problem of institutional change and “evolutionary universals.” A particularly important contribution is to show that the bunching of non-best-response play works as an equilibrium selection device and thus provides a causal mechanism accounting for the evolutionary success of efficient and egalitarian

institutions.

However, taking account of differences in group size and the intentional nature of collective action suggests that the standard stochastic evolutionary game theory model may need further development to be relevant to the historical evolution of institutions. The extensions I have introduced are four. First, non-best response play is intentional rather than accidental. Second, the rate at which non best response takes place is substantial (rather than vanishingly small). Third, non-best response play takes the form of collective action rather than uncorrelated deviant individual behaviors. Four, population sub-groups differ in size, with the less well off typically outnumbering the well off.

I have suggested two reasons why evolutionarily successful institutions may be neither efficient nor egalitarian. First, independently of group size, moderate levels of inequality may deter collective action because the degree of inequality is insufficient to motivate participation. Thus unequal conventions may persist indefinitely. Second, independently of the problem of motivating collective action, the system will spend most of the time in the unequal shares convention because the B's, who prefer this convention, are relatively few in number, so that the likelihood that a random draw will yield a number of them sufficient to displace the convention which they do not prefer is greater than for the A's. This advantage of small numbers is unrelated to conventional reasoning proposed by Olson (1965) and others as to why collective action in large groups is difficult to sustain. The conclusion is that societal inequality of the type described is capable of sustaining highly unequal and inefficient conventions over long periods.

A concern about the stochastic evolutionary game framework is that it applies only to the very long run. For reasonable updating processes, group sizes, and rates of idiosyncratic play, the average waiting times for transitions from one basin of attraction to another are extraordinarily long, certainly surpassing historically relevant time spans, and for some not unrealistic cases exceeding the time elapsed since the emergence of anatomically modern human life. Figure 12 gives the expected number of periods before a transition from an unequal Alternative contract to the Benchmark when the latter is a stochastically stable state.¹⁶ The dynamic assumed is the degenerate case of collective action (whenever there are more than the critical number of A's called to the meeting, they refuse the conventional contract and a transition occurs). Note that, as one would expect, the larger is the number of A's the longer is the waiting time, and that the when the Alternative is as efficient as the Benchmark (the right hand *brs*), it is very persistent even when there are as few as 12 A's. If there are 32 A's, the unequal convention that is only half as efficient as the stochastically stable state persists for an expected one million periods.

¹⁶ I have not explored the waiting times implied by an overlapping generations model ($\omega < 1$). In this case, for a given ϵ the frequency of non best response switches in behavior per period are fewer (in expectations, $\omega\epsilon$ rather than ϵ) but the idiosyncratic play in one period persists over many periods.

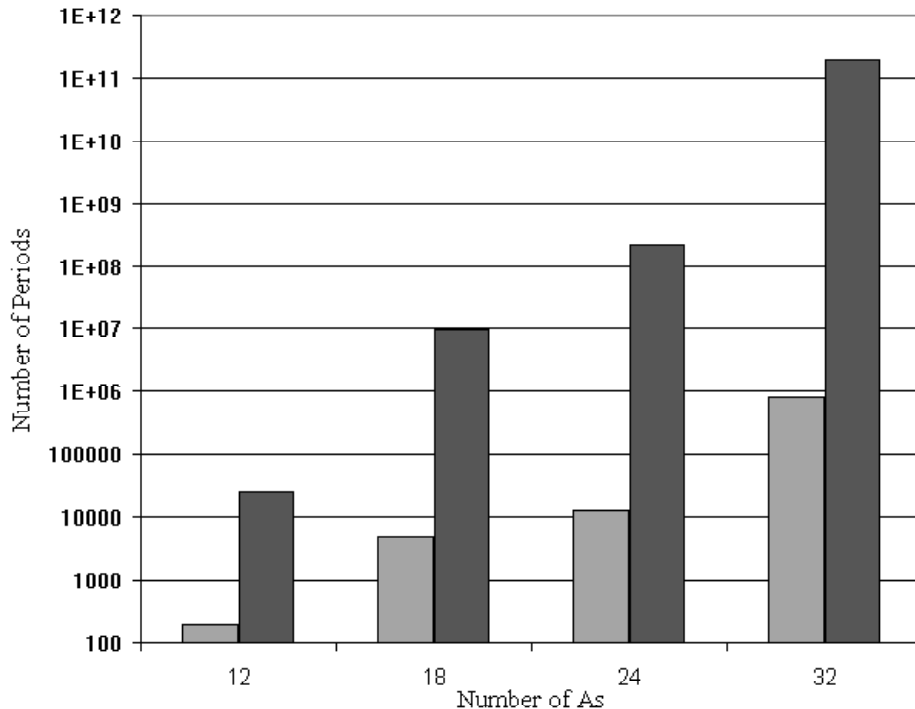


Figure 12 Expected waiting time (in periods, ln scale) for a transition from the alternative to the benchmark convention when the benchmark is a stochastically stable state. The waiting time for this transition depends only on the number of A's. Lefthand bars are for an Alternative with $\sigma = 0.3$ and $\rho = 1$, while the righthand bars are for $\sigma = 0.3$ and $\rho = 2$.

While the biological processes underlying the dynamic referred to in the headquote by Sewall Wright may work over hundreds of thousands of generations, an analogous approach in the social sciences must be relevant to vastly shorter time scales. If the “period” were very short, say, a day, the long waiting times in the figure would be of little concern, but the appropriate period here is an opportunity for collective action to change a convention, and for this, a year or a decade might be more appropriate. Moreover many human groups are larger than those illustrated in the figure, with waiting times correspondingly longer.

However, a number of plausible modifications in the updating process can dramatically accelerate the dynamic process, yielding transitions over historically relevant time scales. Among these are the following.

First, most populations (nations, ethno-linguistic units) are composed of smaller groups of frequently interacting members. Small group membership increases the relative importance of unlikely random events and hence the likelihood that non-best-response play will induce transition

times among conventions at the group level. Because transitions to stochastically stable states are likely to be sustained over long periods, the entire population is likely to transit to the stochastically stable state (all groups eventually making the switch over a relatively short period). Migration among groups or emulation across groups can induce even more rapid transition times for the population as a whole. Hobsbawm and Rude (1968) describe the spread of late 18th and early 19th century Luddite machine-wrecking in England by a process of propagation in small groups and infection of adjacent groups. Because groups are of quite variable size, the process may be considerably accelerated because the transition times will depend not on the mean group size but on the size of the smallest groups.

Second, chance events affect the payoff structures as well as the behaviors of the members of the population. Recall that the location of the internal unstable equilibrium (the saddle, z) and the boundary between the two basins of attraction in Figure 3 is determined by the payoff matrix (equation 1). Variations in environmental effects on payoffs will thus shift the boundary of the basins of attraction, occasionally greatly reducing the size of the basin of attraction of the status quo convention. These effects in conjunction with non-best response play (whether intentional or stochastic) will accelerate the process of transition.

Third, there are generally far more than two feasible conventions, and some of them may be adjacent (that is, the reduced resistances among them are small.) Sewall Wright (1935):263, introducing the passage appearing in the head quote, observed that on a fitness landscape “there is in general a very large number of separate peaks separated by shallow 'saddles'.” A population may traverse a large portion of the state space by means of a series of transitions among adjacent conventions.

Fourth, conformism will reduce the amount of idiosyncratic play. But it also gives rise to positively correlated deviant behaviors -- each member of the population is more likely to adopt a non-best response the more others are doing the same. This produces greater bunching of idiosyncratic play and hence under plausible conditions accelerates the process of transition.

Finally, suppose that individuals may be in two states – active and passive -- that determine whether they necessarily best respond, or may best respond or adopt some other strategy. The passive individuals play a best response (with beliefs formed by the previous period’s distribution of play). Active individuals may engage in non best response behavior depending on the (known) number of other individuals in their class who are active this period. They go to a “meeting” each period and engage in the collective action deliberations described above. Suppose individuals get “activated” with probability ϵ in every period and they stay activated in subsequent periods with probability r (or alternately that they get de-activated with probability r).

In this formulation activation is a heritable mutation that only has phenotypic effects under conditions in which tipping is possible, so there is no selection operating against active individuals until they act in a non best response way. Like Walter Fontana’s *neutral networks* and Motoo Kimura’s *neutral mutations*, this treatment of the activation process combined with the collective action model greatly accelerates the movement to the boundary of the basin of attraction of the status

quo convention The reason is that activation of agents cumulates by drift-like processes undeterred by any opposing selection pressures. Stadler, Stadler, Wagner, and Fontana (2001) Kimura (1968)) Like Timur Kuran's *falsified preferences*, these innovations that may not entail phenotypic changes allow the potential for a tipping event to grow even when there are no behavioral manifestations of this change (Kuran (1995).)

We do not know, of course, whether these modifications of the dynamic modeled here can provide a plausible account of historically observed processes of institutional change. This is an empirical question, which has yet to be explored systematically.

Institutions differ in evolutionarily relevant ways not captured by measures of efficiency, distributional shares, and relative group size, of course. Some institutions may facilitate collective action of the disadvantaged, while others make it more difficult to coordinate. In many situations the effective size of a sub-population may be greatly reduced if it is composed of smaller groups (families, union locals, corporate bodies) that almost always act in unison. Marx, and many since, have believed that the social conditions of industrial capitalism constituted a schoolhouse of revolution, by contrast with earlier institutions of sharecropping, tax farming in societies of independent peasants, and slavery, for example. Barrington Moore (1966) and others, with perhaps greater accuracy, have seen patron-client relationships in agrarian societies and highly unequal systems of land holding as especially vulnerable to revolutionary overturns. These extensions of the basic model may be represented in the differing net benefits of collective action, δ , subscripted by the conventions to which they apply.

Works cited

- Axelrod, Robert and William D. Hamilton. 1981. "The Evolution of Cooperation." *Science*, 211, pp. 1390-96.
- Bergin, James and Barton L. Lipman. 1996. "Evolution With State-Dependent Mutations." *Econometrica*, 64:4, pp. 943-56.
- Boehm, Christopher. 1993. "Egalitarian Behavior and Reverse Dominance Hierarchy." *Current Anthropology*, 34:3, pp. 227-54.
- Bowles, Samuel. 2004. *Microeconomics: Behavior, Institutions, and Evolution*. Princeton: Princeton University Press.
- Bowles, Samuel and Jung-Kyoo Choi. 2003. "The co-evolution of love and hate." *Santa Fe Institute*.
- Bowles, Samuel, Jung-Kyoo Choi, and Astrid Hopfensitz. 2003. "The coevolution of individual behaviors and group level institutions." *Journal of Theoretical Biology*, 223:2, pp. 135-47.
- Foster, Dean and H. Peyton Young. 1991. "Cooperation in the Short and in the Long Run." *Games and Economic Behavior*, 3, pp. 145-56.
- Hobsbawm, Eric and George Rude. 1968. *Captain Swing*. New York: Pantheon.
- Kandori, M. G., G. Mailath, and R. Rob. 1993. "Learning, Mutation, and Long Run Equilibria in Games." *Econometrica*, 61:1, pp. 29-56.
- Kimura, M. 1968. "Evolutionary Rate at the Molecular Level." *Nature*, 217, pp. 624-26.
- Knauff, Bruce M. 1991. "Violence and Sociality in Human Evolution." *Current Anthropology*, 32:4, pp. 391-428.
- Kuran, Timur. 1995. *Private Truths, Public Lies: the Social Consequences of Preference Falsification*. Cambridge: Harvard University Press.
- Marx, Karl. 1963. *The Eighteenth Brumaire of Louis Bonaparte*. New York: International Publishers (original publication date, 1852).
- Moore, Barrington Jr. 1966. *The Social Bases of Dictatorship and Democracy*.
- Moore, Barrington Jr. 1978. *Injustice: The Social Bases of Obedience and Revolt*. White Plains: M.E.Sharpe.
- Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*.

Cambridge, MA: Harvard University Press.

Padover, Saul ed. 1979. *The Letters of Karl Marx*. Englewood Cliffs: Prentice-Hall.

Parsons, Talcott. 1964. "Evolutionary Universals in Society." *American Sociological Review*, 29:3, pp. 339-57.

Smith, John Maynard and G. R. Price. 1973. "The Logic of Animal Conflict." *Nature*, 246, pp. 15-18.

Stadler, Barbel M R, Peter F Stadler, Gunter P Wagner, and Walter Fontana. 2001. "The Topology of the Possible: Formal Spaces Underlying Patterns of Evolutionary Change." *Journal of Theoretical Biology*, 213, pp. 241-74.

Wright, Sewall. 1935. "Evolution in Populations in Approximate Equilibrium." *Journal of Genetics*, 30, pp. 257-66.

Young, H. Peyton. 1998. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton, NJ: Princeton University Press.