Center for Research on Economic and Social Theory
CREST Working Paper

Social Contract III: Evolution and Utilitarianism

Ken Binmore

August, 1988
89-05

DEPARTMENT OF ECONOMICS
University of Michigan
Ann Arbor, Michigan 48109

EVOLUTION AND UTILITARIANISM:

social contract III

by Ken Binmore
Economics Department
University of Michigan
Ann Arbor, MI 48109.

Abstract

The paper takes the simplest possible bargaining game as a paradigm for the coordination problem—i.e. the problem of selecting an equilibrium when many are available. The aim is to explore the circumstances under which evolution will lead to a utilitarian conclusion.

EVOLUTION AND UTILITARIANISM:

social contract III

He who understands *baboon* would do more
toward metaphysics than John Locke.

Charles Darwin.

1. Introduction. This is the third of several free-standing papers whose beginnings lie in Rawls' [1958,1968,*1972*] theory of the social contract. The aim of the sequence of papers is to defend a version of Rawls' "egalitarian"[1] conclusion for a world in which agents are assumed to be constrained only by rational self-interest.

The program entails a very substantial re-evaluation of Rawls' approach, the underlying philosophical attitude of Kant being replaced by a bowdlerized Humean perspective. Among other unusual features, *natural law*[2] is assigned the same status as would seem natural to a natural scientist. Man's capacity to act as a social animal is attributed to his evolutionary history, rather than to transcendental ethical imperatives. The general tone is set by Gibbard [1982], who also draws attention to the importance of *bargaining* in these matters.

Nash's [1950] demand game is perhaps the simplest bargaining model of any interest. Two bargainers make simultaneous demands. If these are compatible, each receives his demand. Otherwise each gets a fixed disagreement payoff. The current paper uses the game as a paradigm for the human "coordination problem"—i.e. the problem of selecting an equilibrium when many are available. The aim is to compare evolution's solution with two man-made alternatives: the utilitarian solution as defended by Harsanyi [1953,1955,1958,*1977*] and the Nash [1950] bargaining solution.

These solution concepts are discussed in parts I and II of the program of papers of which this is part III. However, neither this material nor other foundational issues

will be taken for granted here. This is partly because I hope to make the work accessible to a wider audience; but mostly because I believe that much confusion in the literature derives from straightforward misunderstandings on matters that ought not to be controversial.

Since the evolutionary history of *homo sapiens* is shrouded in mystery, he will be replaced by an idealized homonid whose evolutionary history can be invented. The extent to which such inventions are interesting will depend, of course, on how distant these are from whatever speculations about *homo sapiens* are believed to be plausible. Two possible species of homonid will be considered. The first is the familiar *homo economicus*. He is studied more closely in part IV. The current paper concentrates on another mythical homonid, *homo behavioralis*. Both species are modeled as stimulus-response machines, but *homo behavioralis* is programmed *directly* with behavior, like a chocolate-dispensing machine.

I have made the virtually tautologous point elsewhere [Binmore, 1987,*1988*] that the equilibrium which gets selected will be a function of the equilibriating process, the *libration*, that selects it. A distinction was made between *evolutive* and *eductive* librations. *Homo behavioralis* gets to equilibrium via an evolutive libration: a slow trial-and-error process in which Nature tries out various strategies and those which confer greater fitness survive at the expense of the others. *Homo economicus* gets to equilibrium via an eductive libration. He *thinks* his way to an equilibrium strategy using a classical tâtonnement argument.

A central theme is that *homo economicus* is able to adapt quickly to changing circumstances but that his societies pay a price in having to live with outcomes which are "second-best" from the welfare point of view. *Homo behavioralis* cannot adapt quickly, but section 5 of the current paper shows that, in a fixed environment, evolution may lead a variety of the species to a utilitarian outcome. Nature's hidden hand therefore engineers a utilitarian's "first-best" result.

Section 5 is the heart of the paper. Preceding sections clear the air and lay the groundwork. Later sections seek to extract some meaning from the conclusion. The plan for subsequent papers is that part IV will continue the current study by concentrating on *homo economicus*. The remaining papers in the planned sequence are more remote from the concerns of the current paper. Part V will contain a Humean reinterpretation of Rawls' social contract theory (as opposed to his Kantian view). Within this interpretation, "egalitarian" conclusions can be defended without recourse to hypotheses that need distress any conservative, no matter however red his neck. (This will seem less surprising when one learns how what it is that gets split equally is defined.) The ideas are closely related to those of Buchanan [1975,1976] and Sugden [1986]. Part VI will relate this work to the literature on cooperative bargaining theory. Finally, there will be a paper with the title "A Liberal Leviathan" which aims to put to rest any suspicions of philosophical heterodoxy.

2. *Homo economicus* and *homo behavioralis* . Evolution has gifted *homo economicus* with great intellectual powers. Give him a decidable mathematical problem and he will solve it, instantaneously and without effort. Where information is concealed from him, he will have no difficulty in assigning probabilities to the events that are uncertain whenever it makes sense to do so.[3] But *homo economicus* is more than a sage: he is also a man of action. Evolution has supplied him with preferences. Given these preferences, his behavior is mechanistic: he assesses the consequences of each available action on the basis of the available information, and then chooses an action which is optimal given his preferences over possible consequences.

Why should evolution generate *homo economicus* when *homo behavioralis* is available? Brain power is, after all, an expensive investment. Moreover, homo economicus is somewhat of a nuisance from Nature's[4] point of view. *Homo behavioralis* can be manipulated directly. His behavior can be tailored precisely to the environment

by providing him with a suitable program. *Homo economicus* can only be manipulated *via his preferences*. Nature cannot get directly at his behavior.

In economics, the "principal-agent" problem involves a principal with certain aims who can only act through an agent who may have aims of his own. The principal therefore seeks to design an incentive scheme which minimizes the distortions resulting from having to delegate to the agent. Nature, as a principal, is blind, but her agent, *homo economicus*, can see his environment. Thus, although Nature *loses* fine control in being unable to modify his behavior directly, she *gains* access to information which would not otherwise have been available to her. If the environment changes sufficiently rapidly, the gains will outweigh the losses. *Homo behavioralis* is as blind as his mistress. She can learn about his environment indirectly by observing which types of *homo behavioralis* are reproductively successful. But this learning process is very slow. In brief, *homo economicus* adapts quickly but *homo behavioralis* does not.

The importance of these issues is multiplied when the environment is social. The human environment then necessarily includes entities which are as complex as an individual's own data-processing equipment—i.e. the human brain must cope with an environment containing other humans. If all men were clones, no extra difficulties would arise. From the evolutionary point of view, only a *one-player* game would be involved. Dominated strategies in the Prisoners' Dilemma would be used and Rousseau's "common will" would be a meaningful notion. But to analyze human societies from such a standpoint would seem to place man in the wrong phylum. To quote Hobbes [1651]:

> . . . Bees and Ants[5] live sociably one with another (which are therefore numbered by Aristotle amongst Politicall creatures) . . . and therefore some man may perhaps desire to know, why Man-kind cannot do the same. To which I answer . . . amongst these creatures, the common good differeth not from the Private.

The fallacy which deduces a "common will" from considerations of individual rationality is discussed in section 3 with a view to drawing a perhaps fine distinction between this approach and that of the utilitarians who join Hobbes in speaking of a "common good"—as in Mill's [1851] *summum bonum*. As argued in section 5, the utilitarian view can be seen as requiring only the notion of a "common interest" rather than that of a "common will" and hence utilitarians do not need to look in another phylum for their model of man.

Evolution will not lead homonids to play dominated strategies in the Prisoners' Dilemma because, when this is modeled as a *two-player* game, such behavior is not *in equilibrium.*[6] A utopian therefore faces a *constrained* optimization problem. His "first-best" may not be available as an equilibrium, and so he may have to make do with something which is only "second-best". A utopia which does not recognize this reality will not survive.

This may seem a bleak conclusion if one persists in regarding the Prisoners' Dilemma as the appropriate paradigm for human interaction. But it is rather the *indefinitely repeated* Prisoners' Dilemma which is appropriate. Axelrod [1984], for example, has popularized the fact that high levels of cooperation can be sustained as equilibria in such games. That is to say, cooperative behavior does *not* depend on individuals' abandoning their "private good" for some mythical "common good".

The point here is that is a mistake to to classify the human dilemma as a "cooperation problem" whose solution depends on eliciting behavior from people that is not in their individual best interests. Systems built on such individual self-sacrifice are inherently unstable, and cannot be expected to survive in the long-run. The human dilemma should be seen as a "cordination problem". That is to say, a human society needs to be modeled as a *many* player game with *many* equilibria. Some of these equilibria may call for behavior that a naive observer might characterize as "selfish". In others, although nobody acts except in their own individual best

interest, high levels of "cooperation" may be sustained as a consequence of built–in "punishment schedules" being prescribed for deviants. The coordination problem is that of selecting a "good" equilibrium from those available, rather than a "bad" one. What is important is that only *stable* templates for society be regarded as feasible.

The Nash demand game will be the setting for the study of this problem in the current paper and its successor. It seems to me premature to proceed to a direct examination of multi-period games. The problem of how an equilibrium gets *selected* is then not easily disentangled from that of how an equilibrium is *sustained*. It is, in any case, far from obvious what is the "right" equilibrium theory for such repeated games. Even for a static game like the Nash demand game, the issue is not entirely trivial.

The Nash demand game has many Nash equilibria and, in speaking of the coordination problem, it is the selection of one of these Nash equilibria which is to be understood. But Nature is not a person who chooses. She provides an equilibriating process by means of which an equilibrium emerges. Game theorists evade modeling the details of such processes by introducing "refinements" of the notion of a Nash equilibrium. In the current paper, Maynard Smith's [1982] biological notion of an evolutionary stable equilibrium will be used as an appropriate notion for a society of specimens of *homo behavioralis*. Section 4 introduces this concept. In the next paper, Aumann's [1987] correlated equilibrium is used to capture the idea of the *social* evolution of a "common understanding" in a society of specimens of *homo economicus*. Neither equilibrium notion is a refinement of a Nash equilibrium in the strict sense. Both are simply closely related ideas.

Evolutionary stability in the Nash demand game is studied in sections 5 and 6 for a range of varieties of *homo behavioralis* exhibiting various degrees of smartness. There is a "smartness window" within which utilitarianism is triumphant. On both sides of this window, it is the Nash bargaining solution which emerges.

3. Prisoners' Dilemma. To forestall some possible misconceptions, this section contains some observations on the familiar Prisoners' Dilemma with the payoff matrix of Figure 1A. Adam's payoffs are in the south-west of each cell, and Eve's in the north-east. The version of the Prisoners' Dilemma illustrated is the special case obtained by taking $V = 6$ and $C = 4$ in Maynard Smith's [1982] Hawk-Dove game shown in figure 1B. This explains the labeling of the cooperative strategy as "dove" and the defecting strategy as "hawk".

Figures 1A, 1B, 1C and 1D here

It is supposedly paradoxical that "hawk" strategically-dominates "dove" in the Prisoners' Dilemma. That is to say, "hawk" always pays more whatever the other player does. But if both players choose "hawk", both get less than they would get by cooperating through the playing of "dove".

A common fallacy seeks to escape the paradox by exploiting the symmetry of the Prisoners' Dilemma. If both players reason in the same way to a determinate conclusion, then it follows that only outcomes on the main diagonal of the payoff matrix are possible. Thus, either *both* play "dove" or *both* play "hawk". In a choice between only these two possibilities, *both* players prefer the former. Hence, so the story goes, cooperation is the rational choice.

The flaw is that an explanation is lacking as to *why* players necessarily reason in the same way. If it is argued that they do so because both reason *rationally* from identical premises, then one is not entitled to use rationality a second time without regard to whether this second use, in making a selection *from* the main diagonal outcomes, is consistent with the first use, in getting the outcome *on* the main diagonal. The rationality cow is like other cows: it cannot be milked indefinitely without running dry. A game theorist will agree that rational players will reason the

same in this game. They will reason the same because they will both reason that they should play their dominating strategy. And that is where the story ends.

The fallacy goes back at least as far as Spinoza and still continues to flourish. Gough [1938] quotes the following passage from Spinoza with approval as an unusual example of Kantian "rigor" in his work!

> What if a man could save himself from the present danger of death by treachery?. . . If reason should recommend that, it would recommend it to all men.

The last sentence restricts attention to the main diagonal, whereupon the treacherous outcome becomes "absurd".

Often the pill is sweetened with suggestions of built-in ethical imperatives or urges. Sometimes "shame" or "lack of self respect" are mentioned (with a view to changing the six in figure 1A to something smaller). But, if such ethical drives are to be proposed as primitives, why not simply suppose that people have an in-built urge to cooperate in the Prisoners' Dilemma and leave it at that? It seems to me essential that any primitive urges that are proposed should come with an evolutionary explanation of their origin.

The intuition behind the fallacy is worth exploring further. In an evolutionary context, Adam and Eve may only *nominally* be the players in a game-theoretic sense. Biologists, for example, think of gene packages as the *actual* players in the "game of life". Animals are seen as puppets used to promote the replication prospects of the gene packages which create them. Thus, in the Prisoner's Dilemma, the *same* actual player, or *actor*, may be occupying the roles of Adam and Eve simultaneously. The Prisoner's Dilemma then becomes a *one-player* game like that illustrated in figure 1C. The payoffs in this one–player game are only notional, being obtained simply by addding the payoffs in the corresponding cells of figure 1A. A single actor is conceived of as controlling both Adam's *and* Eve's strategy choice. He is therefore free to select

any cell of figure 1A. His payoff from selecting (hawk, dove), for example, is notionally $0 + 6 = 6$.

The single actor will not necessarily choose an outcome on the main diagonal in this game (definitely not if the sixes in figure 1A are replaced by sixteens). To secure an outcome on the main diagonal, it is necessary that the actor be constrained to play the role of Adam precisely as he plays the role of Eve. Or, if one chooses to think of Adam and Eve as automata, then the actor must be forced to program them identically. Notice that there will be many situations in which such a constraint on the actor makes good sense. The payoff structure of the Prisoner's Dilemma is *symmetric*. Any asymmetries (such as that involved in naming the nominal players Adam and Eve) are therefore arbitrary conventions grafted onto the essential game structure. If no such arbitrary conventions have been established, Adam and Eve will necessarily be döppelgangers: the actor will be unable to distinguish between them and will therefore have no choice but to program them identically.

Having forced symmetry on the actor, the result is a situation in which the intuition behind the fallacy is rescued, but at a very high cost. The players are no longer recognizably human. Instead, they are ant-like representatives of a single "common will".

If the payoffs genuinely measure the fitness of the players, then it seems to me transparent that cooperation is not a rational option for humans, as opposed to ants, in the Prisoners' Dilemma. The consequences are unpleasant only if one follows the fashionable practice of regarding the Prisoners' Dilemma as an ultimate paradigm of the human cooperation problem. But the intractable, one-shot Prisoners' Dilemma occurs only very rarely in real-life.[7] It is rather the indefinitely repeated Prisoners' Dilemma which is the appropriate paradigm. This section has therefore been concerned with a wrong analysis of the wrong game.

The indefinitely repeated Prisoners' Dilemma has many equilibria, some of which generate behavior which is indistinguishable to an observer from open-hearted cooperation [Axelrod,1984]. The real problem for an analyst is *not* whether cooperation can be sustained at all by rational players. Any equilibrium[8] is a possible outcome for rational players. The problem is that of *coordinating* on one of the *many* equilibria normally available.

4. Equilibria. This section distinguishes between Nash equilibria and evolutionarily stable equilibria [Maynard Smith, 1982]. A Nash equilibrium occurs when each player chooses a strategy which is optimal given the strategies chosen by the other players. Thus the strategy-pair (hawk,hawk) is a Nash equilibrium for the Prisoners' Dilemma, but (dove,dove) is not. In the version of the Hawk-Dove game of figure 1D, (hawk,dove) and (dove,hawk) are Nash equilibria. There is also a symmetric Nash equlibrium which requires both Adam and Eve to use the *mixed strategy* of choosing each of hawk and dove with probability 1/2. This makes the opponent indifferent between hawk and dove, and so anything is optimal for him or her including playing each of hawk and dove with probability 1/2.

Evolutionarily stable equilibria require more introduction. Adam and Eve need to be thought of as *nominal* players: specimens of *homo behavioralis* whose behavior is determined by the gene package which is responsible for the design of their stimulus-response pattern. The *actual* players are these gene packages. There is a *normal* gene package and a *mutant* gene package. Adam and Eve are chosen at random from a very large population of nominal players, each of which is controlled either by a normal actor or by a mutant actor.

If *all* the population were controlled by the normal actor, then the set-up would reduce to a one-player situation as studied for the Prisoners' Dilemma in section 2. The mutant actor is introduced to test the *stability* of the one-player set-up. When

will the mutant *invade* the population by expanding whatever small fraction of nominal players it initially controls?

Figure 2A summarizes the necessary information. The average fitness[9] of a nominal player in a Hawk-Dove game who uses strategy s when the nominal opponent uses strategy t is denoted by $f(s,t)$. (Recall that the Hawk-Dove game is symmetric.) The mutant actor instructs the nominal players he controls to use the strategy h: the normal actor instructs those he controls to use strategy d. The initial fraction of the population controlled by the mutant actor is $\pi > 0$.

A normal in the role of Adam expects a payoff of $(1-\pi)f(d,d) + \pi f(d,h)$: a mutant expects $(1-\pi)f(h,d) + \pi f(h,h)$. The payoffs with Eve replacing Adam are the same because of the symmetry of the game. Thus normals will do better than mutants if and only if

$$(1-\pi)f(d,d) + \pi f(d,h) > (1-\pi)f(h,d) + \pi f(h,h). \tag{3.1}$$

Figures 2A, 2B, 2C, and 2D here

For a fully normal population to be invulnerable to such invasion by mutants, (3.1) must be valid for arbitrarily small $\pi > 0$. Imposing this requirement gives the necessary and sufficient conditions for an *evolutionary stable equilibrium*:

$$\begin{aligned} &(i)\ f(d,d) > f(h,d) \\ &\text{OR } (ii)\ f(d,d) = f(h,d) \text{ AND } f(d,h) > f(h,h). \end{aligned} \tag{3.2}$$

The corresponding condition for (d,d) to be a *Nash equilibrium* in the game of figure 2A is simply that

$$f(d,d) \geq f(h,d). \tag{3.3}$$

The immediate point is that the two equilibrium ideas are very close. A mixed strategy pair is an evolutionarily stable equilibrium only if it is a symmetric Nash

equilibrium. In particular, (dove, dove) in the Prisoners' Dilemma is *not* evolutionarily stable. If *homo behavioralis* came equipped with a cooperatively minded "general will", he would soon be extinct.[10]

Two further points need to be made before moving on to the more interesting case of evolutionary stability in games with many Nash equilibria. Notice first that it matters that Adam and Eve are drawn from the *same* population. If Adam and Eve are drawn from *different* populations which evolve *separately*, a mutation appearing in only one of the two populations at a time, then evolutionary stability simply reduces to the requirement

$$f(d,d) > f(h,d) \tag{3.4}$$

i.e. that (d,d) is a "strong" Nash equilibrium.

The second point to notice is that the mutant and normal players were forced to treat Adam and Eve *symmetrically* in the current section. Its results are therefore directly relevant to the discussion of section 3. In the following section, it will be necessary to abandon this symmetry constraint since the results are only of interest when the game under study is *asymmetric*. Difficulties then exist in applying the notion of evolutionary stability as formulated above because there may be no natural way of relating the behavior of a mutant in the role of Adam with that of the same mutant in the role of Eve [Selten, 1980 ; Samuelson, 1988]. Where it is sensible to consider mutants whose behavior in the role of Adam is *independent* of their behavior in the role of Eve, then one is essentially back in the case in which Adam and Eve are drawn from populations that evolve separately, and hence evolutionarily stable equilibrium reduces to strong Nash equilibrium as characterized by (3.4). The latter consideration allows some of the material of the following section to be short-circuited, but only at the expense of some loss of insight into the process by means of which equilibrium is achieved.

5. *Homo behavioralis* and utilitarianism. In the Nash demand game, Adam and Eve make simultaneous demands, x and y. If these are compatible, each receives his demand. Otherwise each receives a disagreement payoff which is normalized to be zero in this paper. One version of the game makes x and y compatible if the pair (x, y) lies in a feasible set[11] X. Sugden [1986] has studied evolutionary stability for this game (which he calls the "division game" and attributes to Schelling) in the case when X is the unit simplex. Unless Adam and Eve are programmed symmetrically, he observes that evolutionary stability does not eliminate any of the many Nash equilibria.

However, in this section, a different variant of the game is studied. The game is nominally to be played by specimens of *homo behavioralis*. Recall that he is a simple stimulus-response machine. He is not equipped to explore his environment scientifically like *homo economicus* and hence cannot exhibit behavior which is tailored to each individual feasible set X. His behavior can only be conditioned on environmental hints and cues that his evolutionary history has exposed as being significant. And the manner in which the species learns from experience is too slow for its behavior to get sharply conditioned on the feasible set X, because this set will never be quite the same the next time that a pair is chosen from the population to act as Adam and Eve. To capture such uncertainty about X, a function $p : \mathbf{R}^2 \to [0, 1]$ is introduced. The number $p(x, y)$ is to be interpreted as the *probability* that the pair (x, y) is compatible.

The Nash demand game was introduced by Nash [1950] as the simplest bargaining model of any interest. Its role in the current paper is to provide a more realistic paradigm of the problem of human cooperation than the Prisoners' Dilemma. If the feasible set X in Nash's demand game is known for certain[12], then the game has many "cooperative equilibria". Each pair of non-negative demands for which (x, y) is Pareto-efficient in X constitutes a Nash equilibrium. The problem is that of

coordinating on one of these. Notice that "non-cooperation" is also available as an equilibrium in the game. All that is necessary is for both players to make demands which are too large to be satisfied, however small the demand of the other.

Nash [1950] observed that the introduction of some shared uncertainty about X limits the Nash equilibria available for selection. The discussion that follows is independent of his conclusions, but it will clarify matters to describe them briefly. (A more extended discussion appears in part II.)

Suppose that $0 < p(x,y) < 1$ if and only if (x,y) is less than a distance of $\epsilon > 0$ from the nearest boundary point of X. For other (x,y), let $p(x,y) = 0$ if (x,y) lies outside X, and let $p(x,y) = 1$ if (x,y) lies inside X. Then the "amount of uncertainty" about X will be small when ϵ is small.

If p is *smooth*, then each Nash equilibrium pair (a,b) of demands lies at a local maximum of the function $\phi:\mathbf{R}^2 \to \mathbf{R}$ defined by

$$\phi(x,y) = xyp(x,y). \tag{4.1}$$

The trivial, or non-cooperative, equilibria (a,b) are those for which $p(a,b) = 0$ and so $(a,b) \notin X$. The cooperative equilibria lie on contours $p(x,y) = \lambda$ with $0 < \lambda < 1$. If these contours adequately approximate the boundary of X when ϵ is small, then all the cooperative Nash equilibria approximate the *Nash bargaining solution* n of X. As illustrated in figure 3A, this is the point of X at which xy is maximized. The *utilitarian solution* u is located where $x + y$ is maximized.[13]

Figures 3A and 3B here

The preceding discussion is relevant because of the close connection between Nash equilibria and evolutionarily stable equilibria as described in section 4. However, the analysis that follows will be phrased directly in terms of a suitable generalization of evolutionarily stable equilibrium. Adam and Eve are taken to be specimens of *homo*

behavioralis drawn from a *single* population with a *common* evolutionary history; not from populations that have evolved separately. This allows proper attention to be paid to the important fact that there will always be a small probability that a mutant will be playing *itself*.

Two cases will be distinguished. In the first case, mutants do *not* recognize each other when they meet. In this case, the Nash bargaining solution n survives as the only possible evolutionarily stable outcome (when ϵ is negligible). In the second case, mutants do recognize each other and, with appropriate auxiliary assumptions, the only possible evolutionarily stable outcome is the utilitarian solution u (when ϵ is negligible).

The results are without interest when the Nash demand game is symmetric, because u and n then coincide. It is therefore necessary to begin by relaxing the requirement of section 4 that the normal and the mutant actors are constrained to treat Adam and Eve identically. Instead, the normal actor programs those of the population he controls to demand a when they find themselves in Adam's role, and to demand b in Eve's role. The corresponding demands for the mutant actor are $x = a + \xi$ and $y = b + \eta$. These two strategies, for the normal and mutant actors, will be denoted by d and h respectively, but the interpretation of these symbols in terms of doves or hawks is now abandoned. The fraction of the population controlled by the mutant will be $\pi > 0$. It is important that each member of the population be as likely to occupy Adam's role as Eve's.

Case 1. Figure 2B displays the payoffs (fitnesses) for Adam and Eve. The results of encounters between the normal and mutant actors are displayed in figure 2C. An actor does not care whether he is replicated through the body of an Adam or Eve and so his payoff is the *average* fitness of an Adam and an Eve under his control. These payoffs are multiplied by two throughout. Consider, for example, the expected

payoff to a normal actor on those occasions when he faces a mutant actor. Half of the time the normal actor will be Adam and the mutant will be Eve. The normal actor plays d and the mutant plays h. The resulting payoff of $ap(a, y)$ can be read off from figure 2B. The other half of the time, the normal actor will be Eve and the mutant will be Adam. Figure 2B then shows that the payoff to the normal actor will be $bp(x, b)$. His expected payoff overall is therefore the average of $ap(a, y)$ and $bp(x, b)$. This appears in figure 2C multiplied by two.

For a normal population to be evolutionarily stable, the effect of the game being played many times between Adams and Eves drawn at random must be to eliminate the mutant foothold in the long run—i.e. the fraction of mutants in the population must diminish to zero. The rate of increase of the mutant group must therefore be less than the rate of increase of the normal group. The criterion is therefore that $M_1 < N_1$, where

$$M_1 = (1-\pi)\{xp(x,b) + yp(a,y)\} + \pi(x+y)p(x,y)$$

and

$$N_1 = (1-\pi)(a+b)p(a,b) + \pi\{ap(a,y) + bp(x,b)\}.$$

It is natural to focus attention on small values of ξ and η. After using Taylor's theorem and suppressing non-linear terms[14] in ξ and η,

$$N_1 - M_1 \approx -\xi(p + ap_x) - \eta(p + bp_y) = -\langle v, \nabla\phi\rangle, \tag{4.2}$$

where the functions are all evaluated at (a, b) and, in the inner product $\langle v, \nabla\phi\rangle$, $v = (\xi/b, \eta/a)$ and ϕ is defined by (4.1).

If $\nabla\phi \neq 0$, ξ and η can be chosen so that $\langle v, \nabla\phi\rangle > 0$. A necessary condition for evolutionary stability is therefore that $\nabla\phi = 0$. Figure 3B illustrates a point (a, b) at which $\nabla\phi(a, b) = 0$. If the contour $p(x, y) = p(a, b)$ adequately approximates the boundary of X, such a point (a, b) approximates the *Nash bargaining solution* n of X.

It is customary to rely on a "hidden hand" to bring about a Pareto-efficient outcome. But here the hidden hand of Nature certainly does not achieve that end. It is true that, when $\epsilon > 0$ is small, n is approximately Pareto-efficient from the point of view of the nominal players Adam and Eve. But the result is *not* efficient from the point of view of the normal actor. At an evolutionarily stable equilibrium, he controls all the population and hence everybody uses strategy d. Thus Adam's payoff is $ap(a,b)$ and Eve's is $bp(a,b)$. The normal actor gets the average of these. If he were unconstrained in his choice of a and b, he would therefore select them so as to maximize $(a+b)p(a,b)$ rather than $abp(a,b)$. That is to say, he would locate at the utilitarian solution (where the expected value of $a+b$ is maximized) and not at the evolutionarily stable equilibrium (where the expected value of ab is maximized).

Although nothing complicated is involved, it is worth expanding on this point since it lies at the heart of what this paper is about. Suppose that a normal population *were* operating the utilitarian optimum at (a,b). A fraction π now become mutants. A mutant thereby gains

$$(1-\pi)\{xp(x,b)+yp(a,y)\}+\pi(x+y)p(x,y)-(a+b)p(a,b)$$
$$=(1-\pi)(M_1-N_1)+\pi\{(x+y)p(x,y)-(a+b)p(a,b)\},$$

if second order terms are neglected. The second term is negative because (a,b) is the utilitarian optimum. When π is small, this term will be outweighed by the first term which will be positive if x and y are suitably chosen. But, if the mutants were to become so numerous that $1-\pi$ became small, then the second term would outweigh the first. That is to say, the long–run gain to becoming a mutant would be negative. The point here is that Nature does not care about ***absolute*** gains. She cares only about *relative* gains.

Case 1 will perhaps serve to clear the air for a study of case 2 which is more promising for utilitarians.

Case 2. Kin-selection is a term which biologists use to describe the process by means of which cooperative behavior can arise between genetically related members of an insider group. It depends on individuals being able to *distinguish* between insiders and outsiders. In essence, insider-insider encounters are treated as one-player games and insider-outsider encounters are treated as two-player games. Axelrod [1984] has emphasized the possibilities that this mechanism has in explaining the "evolution of cooperation".

In case 2, the mutant actor will be allowed to recognize himself when occupying both the role of Adam and the role of Eve. He plays strategy h only when playing himself. When playing against the normal actor, he uses the same strategy d as the normal actor uses all the time.

Adapting the notation of case 1 and using figure 2D,

$$M_2 = (1-\pi)(a+b)p(a,b) + \pi(x+y)p(x,y)$$

and

$$N_2 = (a+b)p(a,b).$$

It follows that

$$N_2 - M_2 = \pi\{(a+b)p(a,b) - (x+y)p(x,y)\} \tag{4.4}$$

and so the normal population will be evolutionarily unstable unless (a,b) is the utilitarian optimum.

An actor will be said to be *blind* if his behavior does not depend on the identity of the actor occupying the opposing role. It has just been established that a blind normal population is vulnerable to invasion by sighted mutants unless it is utilitarian. However, case 1 demonstrates that a blind utilitarian population is vulnerable to invasion by blind mutants. For the utilitarian optimum to be evolutionarily stable, it is therefore necessary that the normal population *not* be blind. This is not an unreasonable supposition. If utilitarianism became established as a consequence of an

invasion by sighted mutants, then the population which remains after the elimination of the original blind dinosaurs will be naturally endowed with the capacity for sight.

A version of this story deserves to be told more carefully. In the beginning there were only blind actors. As described in case 1, evolution generated a blind normal population playing the Nash solution—i.e. using a demand pair (a, b) which maximizes $\phi(a, b) = abp(a, b)$. (Recall that the pair (a, b) is a Nash equilibrium in the Nash demand game and approximates the Nash bargaining solution when ϵ is small.) Nature now produces a sighted mutant. As in case 2, he plays Nash against outsiders, but plays like a utilitarian against himself. The blind population disappears, being unable to compete. The population that remains, the *new* normal population, is now established at the utilitarian outcome. But its members are sighted. They will play Nash against any deviant outsiders that may appear. Such a population is evolutionary stable.

The best that an invading mutant can do when playing against a normal actor (of the new type) who recognizes the mutant as an outsider is to reciprocate the normal actor's play of Nash. When playing against himself, the best the mutant can do is to play like a utilitarian. The optimal invading mutant therefore simply mimics the strategy of the normal actor but is at a disadvantage because he begins with a smaller pool of insiders.

The last sentence seems to me to be important in shedding light on the utilitarian view of a human society as embodying a single "common interest", rather than being directed by a single "common will". All potential actors behave the same because it is in their interests to do so, *not* because they have no other choice. A natural defense of utilitarianism is therefore possible which is free of the fallacy of section 3.

6. *Homo behavioralis* and the serpent. In the preceeding section *homo behavioralis*, like Adam and Eve in the Garden of Eden, does not deceive and is not deceived. What actors believe they see is always true. But is such a state of sublime innocence realistic?

The immediate point is that the story of case 2 in the preceding section requires that sighted actors be able to distinguish without fail between insiders and outsiders. In an ideal world, perhaps nominal players would come with their affiliation written on their foreheads. However, *homo behavioralis* does not inhabit such a world. How is he even to identify an identical twin as such without the risk of error? One may respond that it does not matter if he sometimes makes a mistake. And it is true that the case 2 argument remains valid when insiders sometimes fail to recognise each other, provided that they never identify an *outsider* as an insider. Even if there is only a small probability of making such an error, things go sadly wrong with the case 2 argument.

To check that there is a difficulty with case 2, consider the stage in the story at which a blind mutant population at the Nash solution is invaded by sighted mutants. In the new version, these mutants suffer mildly from myopia. Although they always identify another mutant correctly, their probability of identifying a normal correctly is $\nu < 1$. Then

$$N_3 - M_3 = (1-\nu)(N_1 - M_1) + \nu(N_2 - M_2), \tag{5.1}$$

where the terms on the right hand side are as in (4.2) and (4.4). Observe that $N_1 - M_1$ is independent of the fraction π of mutants, but $N_2 - M_2$ is proportional to π. Thus, no matter how small $1 - \nu$ may be, the first term on the right hand side of (5.1) will dominate the second when π is sufficiently small. It follows that, if normals are sometimes mistaken for mutants, the case 2 story fails. In particular, a

blind population at the Nash solution is *invulnerable* to invasion by slightly myopic mutants.

When studying kin-selection, biologists have no need to be troubled by this conclusion. The reason is that it will seldom be realistic for them to be studying situations in which each pair in a very large population is *equally likely* to be chosen as the pair to play the game. Instead, it will be more likely that Adam and Eve are geographical neighbors. Since a neighbor is much more likely to be kin than a stranger, this changes the story. The single probability π must then be replaced by a probability π_N for normals and a probability π_M for mutants. In particular, π_M, the probability that my opponent is a mutant *given that I am a mutant,* will not necessarily be negligible and hence the second term on the right hand side of (5.1) may dominate. This provides a mechanism through which cooperative behavior can diffuse through a society. Initially small subsocieties become infected with altruistic behavior. These subsocieties thrive at the expense of their selfishly organized rivals and spread the behavior by colonizing the available environment. Such a story would seem to explain, for example, why African hunting dogs regurgitate food for a hungry fellow pack-member.

The evolution of "neighborly ethics" is therefore perhaps not too difficult to understand. But it is the extension of neighborly behavior to *strangers* from another pack which seems to be the interesting question for a social contract study. This is not to deny that much of human behavior between strangers must first have evolved within kin-groups. But it is a separate question whether such behavior is evolutionarily stable in the wider context of interaction across kin-groups.

Choosing a model in which any pair in a very large population is equally likely to be playing captures the idea that it is interaction between strangers that is at issue. The mere fact that Adam is playing Eve then supplies him with no useful information about whether Eve is or is not his neighbor. To identify Eve, Adam

must use some feature exhibited by Eve *herself* rather than being able to rely on the location in which the game is played.

Can Nature help *homo behavioralis* with this identification problem? One thing she can do is to augment whatever characteristic property of nominal players that actors use to make identifications. *In extremis,* she might even print a nominal player's genetic code on his forehead (Howard [1988]). Assuming no mistakes are made in reading such a signal, the difficulties are eliminated, ***provided that*** Nature always tells the truth. But will she?

Consider case 2 at the final stage, with a sighted utilitarian population who play Nash against deviants. But, if deviants are recognized only by the fact that they do not have NORMAL written on their foreheads, then all Nature has to do to produce a successful free-rider is to write NORMAL on his forehead.

Animals often signal to each other with ritual displays. However, according to Maynard Smith [1984], these displays are seldom truthful indicators of an animal's future intentions, nor does the opposing animal perceive them as such.[15] It seems that Nature has little difficulty in generating mutants who tell whatever lie is effective.

In game-theoretic terms, the discussion has turned to the question of *commitment.* A commitment is a *binding* contract made by a player with himself. If commitments can be made and successively advertized as such, then no difficulty exists in rendering threats and promises credible. Without commitment, the credibility of threats and promises needs to be *explained.* Game theory can provide such explanations when the same two players interact over an extended period. Thus, in the indefinitely repeated Prisoners' Dilemma, the TIT-FOR-TAT strategy in Axelrod's [1984] Olympiad simultaneously incorporates a credible threat and a credible promise which opponents can easily learn by observing the actions taken by a player using the strategy. However, in the context of the current paper, it would be a mistake to be diverted into a discussion of these questions. Selecting an

equilibrium in the Nash demand game should be seen as comparable to selecting an equilibrium in the indefinitely repeated Prisoners' Dilemma *as a whole*. The issue is not that TIT-FOR-TAT selects the play of "dove" in each repetition of the Prisoners' Dilemma, but how two strangers managed to select the TIT-FOR-TAT equilibrium[16] in the first place, as opposed, for example, to the HAWK equilibrium (in which "hawk" is always played no matter what).

Making commitments is beyond the capacity of *homo economicus*. He is programmed to optimize. When it comes to honoring what he claimed in the past to be a commitment, he will reconsider and do his duty only if his estimate of the future consequences make it seem optimal to do so. The best he can do is to send costly signals[17]—i.e. take actions which make it expensive for him to back down from the signalled behavior. But such signals are seldom easy to find as Schelling [1960] has made clear with many aptly chosen examples. In contrast, *homo behavioralis* can make commitments. This is not meant in the trivial sense that, because he cannot consider, he certainly cannot *reconsider*. The point is that *homo behavioralis* may make *de facto* commitments simply because a dishonorable mutant has not yet had time to evolve.

The following passage from Kant [1785] may help to illuminate this point:

> In the natural constitution of an organic being . . . let us take it as a principle that in it no organ is to be found for any end unless it is also the most appropriate to that end and the best fitted for it. Suppose now that for a being possessed of reason and a will the real purpose of nature were his *preservation*. . . In that case nature would have hit on a very bad arrangement of choosing reason in the creature to carry out this purpose. For all the actions he has to perform with this end in view. . . could have been maintained far more surely by instinct than it ever can be by reason.

If the principle of the first sentence were valid, the conclusion would follow. But, unless one wants to follow Kant in assigning a teleological role to Nature, it

must presumably be admitted that the principle is *wrong*. Because evolution would lead to a certain outcome if it operated for long enough, it does not follow that it has led to that outcome. As Dawkins [1986] explains, although Nature has unimaginable amounts of time to wander through gene space, gene space is too large for all but a few of its points to be visited.

The temptation to divest *homo behavioralis* of his capacity to make commitments because such a capacity would eventually be evolved away should therefore be resisted. This would be to misunderstand his role as a foil for *homo economicus*. As Kant observes, the latter would be a sorry substitute for a non-sentient homonid equipped with an internal library, costless to consult, which itemized the optimally fit response to *all* possible stimuli. *Homo economicus* owes his metaphorical existence, and *homo sapiens* his real existence, to the fact that they have not had to compete with such a superman. Only varieties of the species *homo behavioralis* of relatively *low* internal complexity are of interest. One of these varieties *is* capable of making a binding promise in that the signals they make do truthfully reflect the behavior they intend. This variety can sustain a utilitarian outcome.

7. Harsanyi's defense of utilitarianism. Rawls [1972] and Harsanyi [1977] seek to follow Kant in making moral recommendations on an *a priori* basis. I believe that Kant is wrong about *a priori* morality for the same reason that he is generally acknowledged to be wrong about *a priori* geometry. Theorems in geometry need to be deduced from axioms and, in order to decide which axioms are interesting and which are not, it is necessary to consult one's experience. Matters are no different in respect of natural laws in ethics. On the other hand, Kant [1785, p56] is clearly right that only a bungler would not seek to distinguish clearly between when he is consulting his experience and when he is making a formal deduction. The current paper is intended to enrich the *experience* which we use in making judgements about which natural laws

are interesting and which are not. The set of natural laws on which this experience bears are those implicitly employed by Harsanyi [1977] in defending utilitarianism.

A reconstruction of his argument in a contractual setting was offered in Part I of the sequence of papers of which this is Part III. For the purposes of the current paper, the details of Harsanyi's argument are not immediately relevant. What is important is that he offers a formal deduction of a version of utilitarianism from certain primitive assumptions of which three deserve special attention. The first concerns the manner by means of which he symmetrizes the problem via the use of what others have called the Harsanyi doctrine. The second concerns how inter–personal comparisons of utilities get made. The third, and crucial, assumption concerns the possibility of making commitments.

The analysis offered in the current paper shows that there is a sense in which each of these assumptions is natural for a variety of *homo behavioralis*. Symmetry appears here as a result of Adam and Eve being equally likely roles for each actual player. Interpersonal comparison problems disappear in the current paper because a payoff is simply an average fitness. The latter are comparable by definition. Commitment is more troublesome. Only with a rather primitive variety of *homo behavioralis* is this defensible. The question is then whether this half-evolved homonid is of any interest as a model of *homo sapiens*.

He is certainly a better model for the behavior of human subjects faced by novel problems in the laboratory than *homo economicus*, whose vaunted intellectual skills seem little in evidence. Learning by trial-and-error seems very much the norm and, while the manner in which motivating ideas get replicated is not biological, the flavor of the process may be much the same. But I do not see that it follows that he is a good model in a social contract discussion. The fact that he is vulnerable to displacement by simple mutations would seem to rule him out entirely for this purpose.

In summary, the suggestion is that the natural laws underlying a utilitarian ethos are recognizably human laws in that they call for the acknowledgement of a "common interest" rather than subjecting individuals to the inflexibilities of a "common will". But the variety of hominid they subsume is too primitive in his behavior to be useful as a model of man as a political animal. It is necessary, albeit reluctantly, to recognize that *homo sapiens* shares too many properties with *homo economicus* for societies whose existence depends on most of their citizens honoring a "common interest" to survive. The best that *homo economicus* can do is to achieve "common understandings" with his fellows about which *equilibrium* should be selected from those available. But, in a society in which groups interact repeatedly over time, a lot can be achieved this way. To quote Hume [1739] on this point:

> . . . I learn to do service to another, without bearing him any real kindness; because I foresee that he will return my service, in expectation of another of the same kind, and in order to maintain the same correspondence of good offices with me or with others. And accordingly, after I have serv'd him and he is in possession of the advantage arising from my action, he is induc'd to perform his part, as foreseeing the consequences of his refusal.

Such a bourgeois view of ethical matters may not be spiritually uplifting, but it does have the advantage of being based on the nature of human beings as-they-are rather than as-we-wish-they-were.

Footnotes

1. "Egalitarian" is not intended in a technical sense. Thus, Rawls' [1972] difference principle is deemed to be egalitarian.

2. It is true, as Hume says, that, since man is a social animal, it is natural that his societies embody codes-of-conduct. But I do not want to follow Hobbes or Hume in calling the codes-of-conduct they propose "natural laws". This gives the impression that the code-of-conduct in question is necessarily universal in all societies.

3. Notice that some care is being taken not to attribute the *impossible* to the variety of *homo economicus* being described. Some varieties of the species are said to be able, for example, to solve mathematically undecidable problems. This creates logical difficulties when equilibria are discussed seriously [Binmore,1987]. A closely related variety of *homo economicus* is the naive Bayesian, who assigns probabilities in circumstances under which it makes no sense to do so [Binmore,1987]. This latter variety of the species is best thought of as extinct.

4. In such assertions, of course, Nature is used in a metaphorical sense. No teleological interpretation is intended.

5. Although not clones, such *hymonoptera* share more genes in common with a sister than they do with a daughter. It is true, of course, that human kin-groups also share genes and that this presumably is important in explaining the behavior of small groups which interact over long periods.

6. Here, of course, it is the *one-shot* Prisoners' Dilemma which is intended. Considerations at each stage of the *repeated* version are not the same. Also, the payoffs are to be understood as measuring the *fitness* of the players.

7. Which is perhaps why laboratory subjects are ill-adapted to playing it.

8. To make this into a tautology, the proviso is added that equilibrium be "suitably" defined.

9. A measure of the number of times the nominal player is likely to replicate the gene package he or she carries.

10. Of course, ants are not extinct but they do not have the same reproductive arrangements as men. Amongst the *hymenoptera* an individual has more genes in common with a sister than with a daughter. Dawkins [1976] attributes the fact that true social insect societies have arisen independently eleven times among the *hymenoptera* and only once elsewhere to this biological phenomonen. The exceptional, and unexplained, case is that of the termites.

11. Usually assumed to be closed, bounded above, comprehensive and convex.

12. So that $p(x,y) = 1$ when $(x,y) \in X$, and $p(x,y) = 0$ when $(x,y) \notin X$.

13. In the terminology of Part II, section 10, this is the utilitarian solution with weights $\omega_A = \omega_E = 1$ and disagreement direction $d = (1,1)$.

14.

$$N_1 - M_1 \approx (1-\pi)\{(a+b)p - ap - \xi p - a\xi p_x - bp - \eta p - b\eta p_y\}$$
$$\pi\{ap + a\eta p_y + bp + b\xi p_x - (a+b)p - (\xi+\eta)p - (a+b)(\xi p_x + \eta p_y)\}$$
$$= -\{(\xi+\eta)p + a\xi p_x + b\eta p_y\}.$$

15. Signals about future intentions must be distinguished from signals about what the payoffs in the game are. The latter can often be signalled in a manner which makes lying impossible and are apparently often observed in effective use. But since the paper deals only with games of complete information, no room exists for such signals.

16. It is a Nash equilibrium for both players to play TIT-FOR-TAT in the *indefinitely* repeated Prisoners' Dilemma provided that the stopping probability after

each repetition is sufficiently small. It is, of course, *not* a Nash equilibrium in the *finitely* repeated Prisoners' Dilemma studied by Axelrod. Nor is it a *subgame-perfect* equilibrium [Selten,1975] in the indefinitely repeated game.

17. For a contrary view, see, for example, Farrell [1987]. Whether one believes that *cheap* signals can be effective depends upon what one believes to be acceptable as an equilibrium definition.

References

R Aumann [1987], " "Correlated equilibrium" as an expression of Bayesian rationality", *Econometrica*, 55, 1-18.

T. Axelrod [1984], *The Evolution of Cooperation*, Basic Books, New York.

K. Binmore [1987/1988], "Modeling rational players, I and II", *Economics and Philosophy*, 3, 179–214.

K. Binmore [1988], "Social Contract, I–VI", University of Michigan, mimeo. "Social Contract I: Harsanyi and Rawls" is to appear in *Economic Journal.*

J. Buchanan [1975], *The Limits of Liberty*, U. of Chicago Press, Chicago.

J. Buchanan [1976], "A Hobbsian interpretation of the Rawlsian difference principle", *Kyklos* 29, 5–25.

R. Dawkins [1976], *The Selfish Gene*, Oxford University Press, Oxford.

R. Dawkins [1986], *The Blind Watchmaker*, Penguin Books, London.

J. Farrell [1987], "Cheap talk, coordination and entry", *Rand Journal of Economics*, 18, 61–94.

A. Gibbard [1982], "Human evolution and the sense of justice", *Midwest Studies in Philosophy* 7, eds., P. French, T. Vehling and H. Wettstein, U. of Minnesota Press, Minneapolis.

J. Gough [1938], *The Social Contract*, Clarendon Press, Oxford.

J. Harsanyi [1953], "Cardinal utility in welfare economics and in the theory of risk-taking", *Journal of Political Economy*, 61, 434–435.

J. Harsanyi [1955], "Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility", *Journal of Political Economy*, 63, 309–321.

J. Harsanyi [1955], "Cardinal utility in welfare economics and in the theory of risk-taking", *Journal of Political Economy,* 63, 309–321.

J. Harsanyi [1958], "Ethics in terms of hypothetical imperatives", *Mind*, 47, 305–316.

J. Harsanyi [1977], *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, Cambridge U. Press, Cambridge.

T. Hobbes [1651], *Leviathan*, ed. C.B. Macpherson, Penguin Classics, London, 1968.

J. Howard [1988], "Cooperation in the Prisoners' Dilemma" *Theory and Decision*, 24.

D. Hume [1739], *A Treatise of Human Nature*, ed. Mossner, Penguin, Harmondsworth, 1969.

I. Kant [1785], *Groundwork of the Metaphysic of Morals*, ed. H. Paton, Harper and Row, New York, 1964.

J. Maynard Smith [1982], *Evolution and the Theory of Games*, Cambridge U. Press, Cambridge.

J.S. Mill [1851], "Utilitarianism", in *Utilitarianism and other Essays*, ed. A. Ryan, Penguin Books, Harmondsworth, 1987.

J. Nash [1950], "The bargaining problem," *Econometrica*, 18, 155–162.

J. Rawls [1958], "Justice as fairness", *Philosophical Review*, 57.

J. Rawls [1968], "Distributive justice, some addenda," *Natural Law Forum*, 13.

J. Rawls [1972], *A Theory of Justice*, Oxford U. Press, Oxford.

J. Maynard Smith [1982], *Evolution and the Theory of Games*, Cambridge U. Press, Cambridge.

L. Samuelson [1988], "Evolutionary stability in asymmetric games", discussion paper, State University of Pennsylvania.

R. Selten [1975], "Re-examination of the perfectness concept for equilibrium in extensive form games," *International Journal of Game Theory* 4, 22–5.

R. Selten [1988], "A note on evolutionarily stable strategies in asymmetric animal conflicts", *Journal of Theoretical Biology* 84, 93–101.

R. Sugden [1986], *The Economics of Rights, Cooperation and Welfare*, Basil Blackwell, Oxford.

		Eve: dove	Eve: hawk
Adam	dove	3, 3	0, 6
	hawk	6, 0	1, 1

figure 1A: Prisoners' Dilemma

		Eve: dove	Eve: hawk
Adam	dove	V/2, V/2	V, 0
	hawk	0, V	(V-C)/2, (V-C)/2

figure 1B: Hawk-Dove Game

	dove	hawk
dove	6	6
hawk	6	2

figure 1C: one-player version

		Eve: dove	Eve: hawk
Adam	dove	1, 1	2, 0
	hawk	0, 2	-1, -1

figure 1D: V=2, C=4

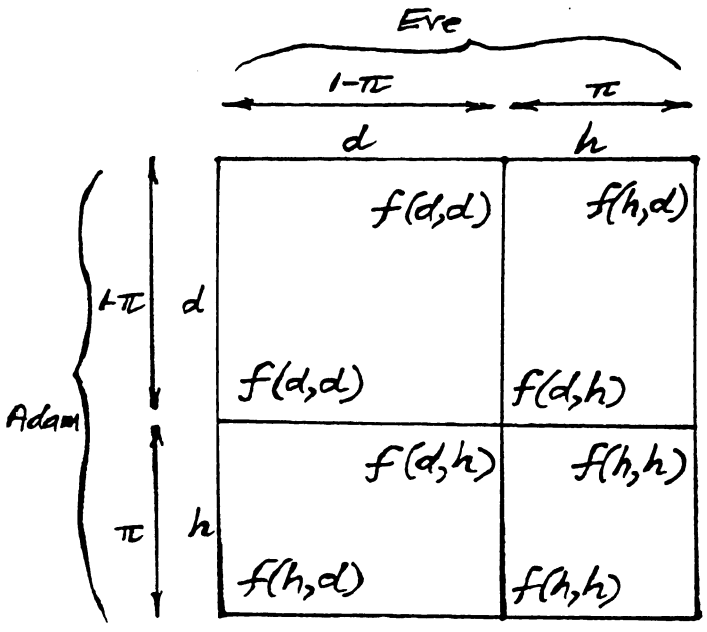

figure 2A

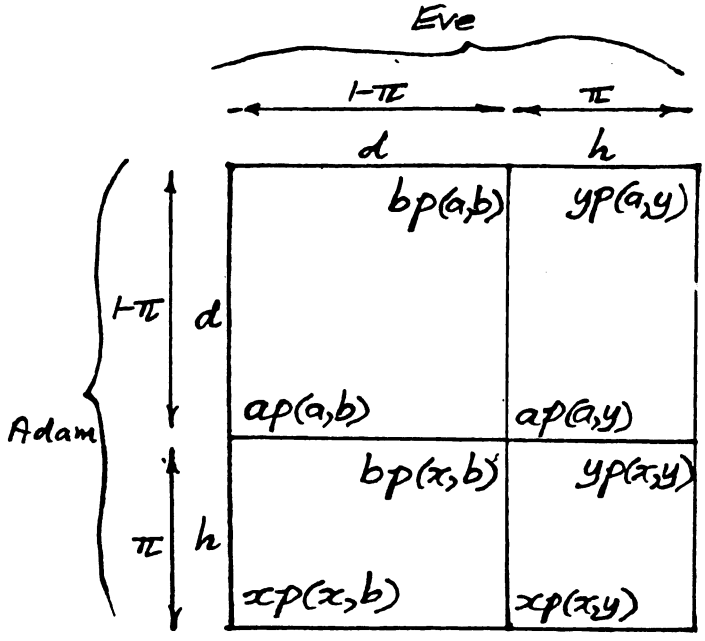

figure 2B

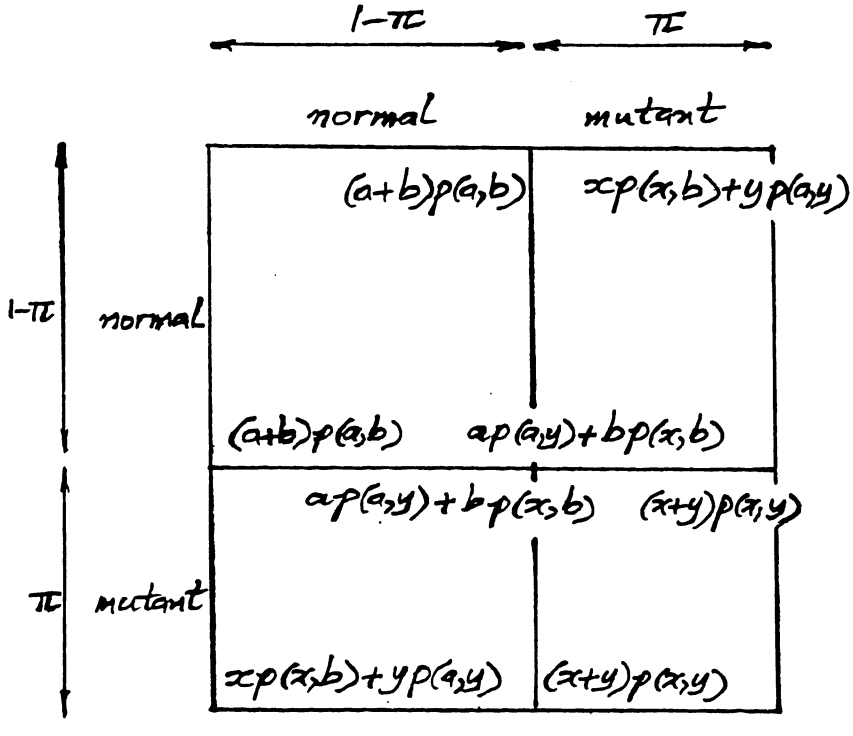

figure 2C

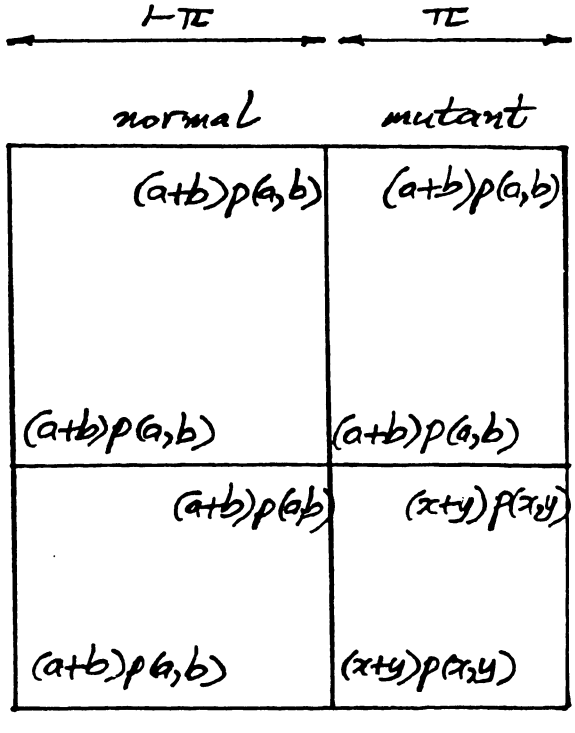

figure 2D

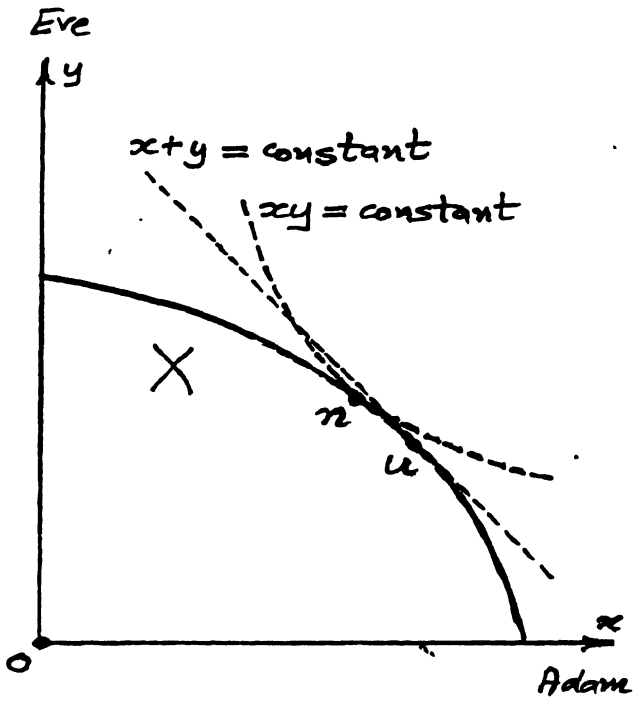

figure 3A

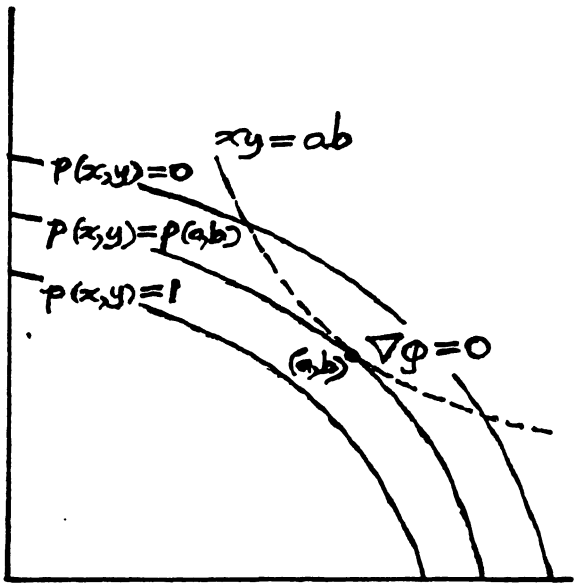

figure 3B

Recent Crest Working Papers

88–1: Carol A. Jones, Suzanne Scotchmer, "The Social Cost of Uniform Regulatory Standards in a Hierarchical Government" December, 1987.

88–2: Ted Bergstrom, Judy Roberts, Dan Rubinfeld, Perry Shapiro, "A Test for Efficiency in the Supply of Public Education" December 12, 1987.

88–3: Mark Bagnoli, J. Bradley Barbeau, "Competition and Product Line Choice" February, 1988.

88–4: Severin Borenstein, Paul N. Courant, "How to Carve a Medical Degree: Human Capital Assets in Divorce Settlements" December, 1987.

88–5: Mark Bagnoli, Stephen W. Salant, Joseph E. Swierzbinski, "Pacman Refutes the Coase Conjecture: Durable-Goods Monopoly with Discrete Demand" January, 1988.

88–6: Jonathan Cave, Stephen W. Salant, "A Median Choice Theorem" December 29, 1987.

88–7: Mark Bagnoli, Naveen Khanna, "Why Are Buyers Represented by Seller's Agents When Buying a House?" December, 1987.

88–8: Mark Bagnoli, Roger Gordon, Barton L. Lipman, "Takeover Bids, Defensive Stock Repurchases, and the Efficient Allocation of Corporate Control" October, 1987.

88–9: Mark Bagnoli, Barton L. Lipman, "Private Provision of Public Goods can be Efficient" November, 1987.

88–10: Michelle J. White, "Urban Commuting Journeys are Not "Wasteful"" February, 1988.

88–11: Avery Katz, "A Note on Optimal Contract Damages When Litigation is Costly" February, 1988.

88–12: Ted Bergstrom, Jeffrey K. MacKie-Mason, "Notes on Peak Load Pricing" February, 1988.

88–13: Jerry A. Hausman, Jeffrey K. MacKie-Mason, "Price Discrimination and Patent Policy" February, 1988.

89–01: Mark Bagnoli, Severin Borenstein, "Carrot and Yardstick Regulation: Enhancing Market Performance with Output Prizes" October, 1988.

89–02: Ted Bergstrom, Jeffrey K. MacKie-Mason, "Some Simple Analytics of Peak-Load Pricing" October, 1988.

89–03: Ken Binmore, "Social Contract I: Harsanyi and Rawls" June, 1988.

89–04: Ken Binmore, "Social Contract II: Gauthier and Nash" June, 1988.

89–05: Ken Binmore, "Social Contract III: Evolution and Utilitarianism" June, 1988.

89–06: Ken Binmore, Adam Brandenburger, "Common Knowledge and Game Theory" July, 1988.

89–07: Jeffrey A. Miron, "A Cross Country Comparison of Seasonal Cycles and Business Cycles" November, 1988.

89–08: Jeffrey A. Miron, "The Founding of the Fed and the Destabilization of the Post-1914 Economy" August, 1988.

89–09: Gerard Gaudet, Stephen W. Salant, "The Profitability of Exogenous Output Contractions: A Comparative Static Analysis with Application to Strikes, Mergers and Export Subsidies" July, 1988.

89–10: Gerard Gaudet, Stephen W. Salant, "Uniqueness of Cournot Equilibrium: New Results from Old Methods" August, 1988.

89–11: Hal R. Varian, "Goodness-of-Fit in Demand Analysis" September, 1988.

89–12: Michelle J. White, "Legal Complexity" October, 1988.

89–13: Michelle J. White, "An Empirical Test of the Efficiency of Liability Rules in Accident Law" November, 1988.

89–14: Carl P. Simon, "Some Fine--Tuning for Dominant Diagonal Matrices" July, 1988.