Center for Research on Economic and Social Theory

CREST Working Paper

Social Contract II:
Gauthier and Nash

Ken Bimore

July, 1988
89-04

DEPARTMENT OF ECONOMICS
University of Michigan
Ann Arbor, Michigan 48109

GAUTHIER AND NASH:

social contract II

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

GAUTHIER AND NASH: social contract II

by Ken Binmore*

> Remember always to study power as it is, not as you would have it be.
>
> Machiavelli.

0. Introduction. This is the second 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.

It should be emphasized that the program entails a very substantial re-evaluation of Rawls' approach. This began in Part I with a refusal to follow Rawls in rejecting orthodox decision theory. Instead, Harsanyi's alternative formulation [1953,1955,1958,*1977*] was appraised. Although his defense of utilitarianism is rejected, Harsanyi's utility theory framework forms the basis of my own approach. However, Harsanyi and Rawls[2] settle bargaining questions by appealing to an implicit "ideal observer" and this will not suffice for my purposes.

This is a view shared with Gauthier [1986]. However, I differ from Gauthier in taking an orthodox view on how bargaining questions should be resolved, and it is this difference that occasions the current paper. But the paper is not a piece of "infighting among bargaining theorists" [Gauthier, 1986, p146]. It is intended as a user's guide to what I see as the mainstream position among contemporary game theorists on bargaining issues. This mainstream has only recently emerged from some

*The material of the paper is an expanded extract from a long ST/ICERD discussion paper "Game Theory and the Social Contract" (88/170). This supercedes an earlier ST/ICERD discussion paper (84/108) with the same title.

notoriously muddy backwaters. Such an exposition may therefore be timely for those who do not follow the economics literature. Some technicalities need to be faced, but nothing that is likely to cause difficulties for anyone with a serious interest in bargaining theory. The chief points to be made concern the manner in which the Nash Bargaining Solution is defended and the circumstances under which its use is appropriate. Other bargaining solutions receive incidental consideration.

Apart from an occasional comment, specific reference to Gauthier's "Morals by Agreement" is left to a concluding section. As for the sequence of papers of which this is part II, the first part contains a reconstruction of Rawls à la Harsanyi. Parts III and IV will be asides on evolutionary issues. Part V will contain a Humean reinterpretation of Rawls' social contract theory (as opposed to his Kantian view). With this reinterpretation, "egalitarian" conclusions can be defended without recourse to hypotheses that need distress any conservative, no matter how 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 material to the bargaining solutions covered here. Finally, there will be a paper with the title "A Liberal Leviathan" which offers a philosophical overview.

1. The Nash Program. The most important point concerns the interpretation of the *Nash Bargaining Solution.* Nash [1950] defended this in two ways. His axiomatic defense is well-known and has been extensively discussed. His second defense depends on the analysis of a specific model of a negotiation process in which both bargainers are restricted to making simultaneous take-it-or-leave-it *demands* (section 3 and section 8(iii) of Part I). With some vestigial uncertainty about precisely what is available, Nash shows that optimal non-cooperative play in this demand game leads to an outcome which approximates his bargaining solution. This paper contends that

Nash's axiomatic defense cannot be properly understood without an appreciation of at least some of the non-cooperative bargaining games of which Nash's demand game is one example. The two lines of attack are simply different aspects of a single conceptual approach nowadays referred to as the *Nash program* [1951,1953].

The idea is very reductionist. Where opportunities for negotiation exist, the proposal is that the various ploys open to the negotiators be modeled as formal moves in a non-cooperative game. The term "non-cooperative" refers to the manner in which the negotiation game is to be analyzed. The players are assumed to be motivated only by rational self-interest, and attention therefore centers on locating *equilibria* of the negotiation game. (Confusion sometimes arises here between a *Nash equilibrium* and the *Nash bargaining solution.* In a two-player game, to which attention is always confined in this paper, a Nash equilibrium is a pair of strategies, one for each player, such that each strategy is an optimal response to the other. The Nash bargaining solution, on the other hand, is a particular rule for selecting a pair of *payoffs* from those available and takes no explicit account of strategic questions.) If one *can* formalize the relevant negotiation procedure adequately and then locate an appropriate equilibrium, whether Nash or some more refined variant, which is unequivocally the "right" equilibrium, then the bargaining problem is solved. The solution of the bargaining problem is the "right" equilibrium outcome—i.e. the payoffs the players receive if they both use their equilibrium strategies during the play of the negotiation game.

Of course, major difficulties exist. (See Binmore/Dasgupta [1987] for a fuller account.) Nash's axiomatic approach is designed to short-circuit these. He asks what properties one might anticipate the equilibrium outcome to have. He then codifies these properties as "axioms" and demonstrates that only one pair of payoffs satisfies the axioms: namely, the Nash bargaining solution. Having arrived at this conclusion. it is tempting to discard the conceptual framework of non-cooperative negotiation games as redundant scaffolding. But this would be a serious mistake. One *must* have a view on what the relevant negotiation game is before one can make a judgement on what properties its equilibrium outcome ought to have. If the negotiation game does not belong to the class of those Nash had in mind when formulating his axioms, then the Nash bargainng solution will be wrong. Even when the Nash bargaining solution is right, it is easy to go astray by applying it wrongly (section 6).

This account of the Nash program has been included largely so that a list can be made of interpretations of the Nash bargaining solution that I specifically want to deny as relevant to my social contract theory. If one hopes to convince a sceptical conservative, there is no point in beginning with hypotheses which he will be bound to reject. It is therefore important to stick with an interpretation in which agents act like *homo economicus*. That is to say, they *optimize as individuals.*

(1) Collective rationality interpretations. Gauthier [1986,p129] asks: "Whether there are principles of rational bargaining with the same context-free universality of application as the principle of expected utility maximization. . . ?" He then follows numerous others in offering some idiosyncratic principles of his own which lead to a version of the "bargaining solution" of Kalai and Smorodinsky [1975]. I think that the answer to his question is "no". Without the context of a negotiation *procedure*, the bargaining problem is *indeterminate.* The Nash program calls for a relevant context to be established and then asks that the appropriate principles of rational bargaining *for that context* be *deduced* from the principle of expected utility maximization. Nash's axioms, notably the "independence of irrelevant alternatives" (section 2), are *not* to be seen as context-free, collective rationality principles.

(2) Behavioral interpretations. Von Neumann and Morgenstern [1944] say that the bargaining problem is outside the scope of game theory because its resolution depends on the *psychology* of the bargainers. Nash [1950] refers to equal bargaining *skill* on the part of the players (although he explicitly corrects himself later [1953]). Harsanyi [1977] quantifies such a notion, in the context of an informal dynamic bargaining model, by reviving an idea of Zeuthen [1930] and elevating this to the status of a principle: *Zeuthen's principle.* More recent advances in game theory mean that it is possible to dispense with such *ad hoc* behavioral assumptions provided that what the players actually can or cannot do is adequately tied down. It is true that a price is paid for this. The game-theoretic analysis for a chimpanzee playing Chess, for example, becomes the same as for Alekhine.[3] If both *optimize*, both will do the same. But this is the price that is always paid for using *homo economicus* as a model for *homo sapiens.* In any case, interpretations of the Nash axioms in terms of the psychology of the players, or in terms of variants of Zeuthen's obsolete principle, are *not* to be admitted.

(3) Ethical interpretations. These are the most invidious. Raiffa [1953] compared a number of axiom systems, of which Nash's was one, as candidates for a "fair" arbitration scheme. Since then, the idea that the Nash bargaining solution is properly to be interpreted in this manner has taken on a life of its own and the Nash bargaining solution is routinely rejected on the grounds that it has no merit as an ethical concept. My own view, which simply echoes what is orthodox in social choice theory, is that ethical decisions require inter-personal comparison of utility units. Such comparisons are expressly ruled out-of-court by Nash's axioms. That is to say, *it is agreed* that the Nash bargaining solution has no merit as an ethical concept. When used, it is to predict what the result would be, under certain ideal circumstances, if agents were to act as individual optimizers. The issue is therefore never whether or not one *likes* the result. As with $2 + 2 = 4$, the only issue is whether the result is accurate.

2. Nash's axioms. As in part I of this sequence of papers, there will be two protagonists: Adam and Eve. If they are rational in their attitudes to risk, in the sense axiomatized by Von Neuman and Morgenstern, their personal preferences can be represented by utility functions $\varphi_A : S \to \mathbf{R}$ and $\varphi_E : S \to \mathbf{R}$ in such a way that Adam and Eve always behave as though maximizing *expected* utility. The set S of feasible deals on which Adam and Eve might agree is a subset of the set S on which their preferences are defined.[4] A particular member s_0 of S_0 represents the current *status quo* (on which more in section 5). In terms of Von Neumann and Morgenstern utilities the *feasible set* is

$$X = \{(\varphi_A(s), \varphi_E(s)) : s \in S_0\}$$

The point $d = (\varphi_A(s_0), \varphi_E(s_0))$ will be called the *disagreement point*. The pair (X, d) is said to be a "Nash bargaining problem".

Fig.1A illustrates a pair (X, d). The diagram incorporates the assumptions about X that will routinely be made: namely, that X is closed, bounded above, convex and comprehensive.[5] For comment on these assumptions, see part I, section 8(iii). It will also be assumed that d is an interior point of X.

Fig.1A also illlustrates the *weighted Nash bargaining solution* n for (X, d) corresponding to the *bargaining powers* $\beta_A > 0$ and $\beta_E > 0$. This is the value of $x = (x_A, x_E)$ at which the weighted "Nash product"

$$\pi = (x_A - d_A)^{\beta_A}(x_E - d_E)^{\beta_E}$$

is maximized subject to the requirements that $x \in X$ and $x > d$. Notice that only the ratio β_A/β_E is significant. The alternative geometric characterization, that $LN/MN = \beta_A/\beta_E$ as illustrated in fig.1A, is often less cumbersome to use.

Figures 1A and 1B here

Two properties of the weighted Nash bargaining solution require emphasis. The first is that the deal that results from its use - i.e. the deal t which satisfies $n = (\varphi_A(t), \varphi_E(t))$, is ***independent of the calibration of the utility scales*** built into φ_A and φ_E. (Nothing changes, for example if x_A and d_A are replaced, in the definition of the Nash product, by $Bx_A + C$ and $Bd_A + C$, where $B > 0$ and C are constants). A bargaining solution which failed to satisfy this requirement would be incoherent from the viewpoint of the Nash program. This is because it is only the players' personal *preferences* which actually have substance. A utility function is a *mathematical invention* whose purpose is simply to ease the task of *describing* an individual's personal preferences. It is the latter which are the *primitives* in Von Neumann and Morgenstern's theory. An individual does ***not*** prefer s to t because $\varphi(s) > \varphi(t)$: on the contrary, $\varphi(s) > \varphi(t)$ ***because*** s is preferred to t. In particular, $B\varphi + C$ represents an individual's preferences just as well as φ.

In the Von Neumann and Morgenstern theory, $B\varphi + C$ and φ are ***indistinguishable***. But an ethical theory needs to distinguish them as a preliminary to a discussion of the inter-personal comparison of utilities (Part I, section 6). Such a theory therefore needs to add *further primitive assumptions* to those of Von Neumann and Morgenstern. There is, of course, a tradition in which utility functions themselves are taken as primitives. Significance is then assigned to utility scales by *fiat.* Such an approach seems to me to beg the question. Bertrand Russell said that Mathematics is the subject where we do not know what we are talking about and do not care whether what we are saying is true. But this surely ought not to be true of Ethics.

The ***independence of irrelevant alternatives*** is illustrated in fig.1B. The feasible set X has been replaced by a smaller feasible set Y which still contains the bargaining solution n for the problem (X, d). The condition requires that the bargaining solution for (Y, d) ***remains*** at n under such circumstances. The following result of Roth [1979] (also paper 1 of Binmore/Dasgupta [1987]) concerns functions f which map each pair (X, d) onto a point $n = f(X, d)$ in the set X. *Strong individual rationality* is an unfortunate way of expressing the very mild requirement that $n > d$.

Proposition. *A function f satisfies the properties*

1. *Strong individual rationality*
2. *Independence of utility calibration*
3. *Independence of irrelevant alternatives*

if and only if $\beta_A > 0$ and $\beta_E > 0$ exist such that $n = f(X, d)$ is the weighted Nash bargaining solution corresponding to β_A and β_E.

Nash [1950] also required *Pareto efficiency* and *symmetry*. The first of these is a *conclusion* in the above formulation. The role of an additional assumption of *symmetry* is simply to ensure that $\beta_A = \beta_E$. This is the case normally understood when the Nash bargaining solution is mentioned.

The first requirement of the proposition is very mild. It has been argued that the second is no constraint at all in the current context. All therefore rests on the much maligned third property.

It may be that "collective rationality" is genuinely meaningful in some situations in spite of Arrow's theorem and other impossibility results.[6] If so, then the *independence of irrelevant alternatives* is certainly one of the candidates for a property that collective rationality should satisfy. But I do not think it should be defended on such grounds in a bargaining context, although its name clearly invites such an interpretation (the alternatives in X but not in Y being "irrelevant", since they are not chosen when n is available).

In game theory jargon, the *independence of irrelevant alternatives* should be seen as an *implementability* requirement[7]. Roughly speaking, a payoff pair is *implementable* if there exists a game (of some specific type) for which the payoff pair is an equilibrium outcome. Here, as elsewhere, the water is muddied by the existence of theorems which are not strictly relevant. These concern *precise* implementation.[8] But one ought only to care about being able to implement within as small a margin of error as one chooses. However, this is a technical point. What needs to be emphasized here is that, if it is indeed true that the *independence of irrelevant alternatives* can sensibly be seen as an implementability requirement, then it is *necessary* that it be insisted upon, given the aims of the Nash program. But is it sensible to see it this way? This is not a question which is easy to answer in the abstract. It seems necessary to look at some specific bargaining games.

3. Nash's demand game. Adam demands x. Simultaneously, Eve demands y. Each gets his or her demand unless these are incompatible, in which case each gets nothing. The probability that a pair (x, y) of demands is compatible is $p(x, y)$. If the demands are Von Neumann and Morgenstern utilities,[9] and (a, b) is a Nash equilibrium, then $x = a$ must maximize $xp(x, b)$ and $y = b$ must maximize $yp(a, y)$. The reason is that $xp(x, b)$ is the *expected* payoff to Adam given that Eve has chosen b. Similarly, $yp(a, y)$ is the *expected* payoff to Eve given that Adam has chosen a.

When differentiation is legitimate, it follows that

$$\left.\begin{aligned} ap_x(a,b) + p(a,b) &= 0 \\ bp_y(a,b) + p(a,b) &= 0 \end{aligned}\right\} \qquad (3.1)$$

If the contour $p(x, y) = p(a, b)$ admits a tangent line at the point (a, b), this is given by

$$p_x(a,b)(x-a) + p_y(a,b)(y-b) = 0$$

From (3.1), it follows that the tangent line is

$$\frac{x}{2a} + \frac{y}{2b} = 1$$

as illustrated in fig.2A. That is to say, (a, b) is the ***Nash bargaining solution***[10] for the problem $(Y, 0)$, where Y is the set shaded in fig.2A.

Figures 2A and 2B here

Of course, when there is no uncertainty about what demands are compatible, then differentiation is ***not*** legitimate because p is discontinuous (since then $p(x, y) = 1$ when $(x, y) \in X$, and $p(x, y) = 0$ when $(x, y) \notin X$). In this case, ***any*** Pareto-efficient (x, y) of X with $x > 0$ and $y > 0$ is a Nash equilibrium. Without a criterion for selecting one of these as a solution of the demand game, one is left therefore with an *indeterminate* situation[11]. To obtain a determinate result, some ***vestigial*** uncertainty is required. It is natural to model this by supposing that each contour $p(x, y) = \lambda$, with $0 < \lambda < 1$, approximates the boundary of X as indicated in fig.2B. Any Nash equilibrium of the demand game with such a dash of uncertainty will then approximate the Nash bargaining solution for the problem $(X, 0)$.

Is it genuinely appropriate to attach such significance to the possible existence of vestigial uncertainties? Evolutionary arguments in support of a positive answer appear in Parts III and IV. Other, more conventional, arguments appear in Binmore/Dasgupta [1987,pp 161–167]. But perhaps it is enough, for the current paper, to observe that nothing is ever certain in life, not even death and taxes.

Return now to the ***independence of irrelevant alternatives*** and consider the following abstract bargaining game based on Nash's demand game. Two players choose negotiation strategies s_1 and s_2 from strategy sets S_1 and S_2, which are fixed independently of X and d. A fixed function $g : S_1 \times S_2 \rightarrow \mathbb{R}^2$ determines the payoffs. The payoff pair is $g(s_1, s_2)$, unless $g(s_1, s_2) \notin X$, in which case the payoff pair is d. Suppose the game always has a ***unique***, non-trivial, Nash equilibrium[12] generating the outcome n. What now happens if X is replaced by Y as in the ***independence of irrelevant alternatives***? Observe that the ***same*** negotiation strategies ***remain*** in equilibrium and so generate the *same* outcome n. In this situation the equilibrium outcome *must* therefore satisfy the *independence of irrelevant alternatives.*

4. Commitment and time. I will not say that it is *obviously* false that there is a "context-free" bargaining solution in the sense of the Nash program. Nevertheless it *is* false. The Nash demand game implements the Nash bargaining solution. On the other hand, Moulin [1984] constructs a bargaining game which implements the bargaining solution of Kalai and Smorodinsky [1975] (see section 12). Does one conclude that both bargaining solutions are equally "legitimate"? I think not. Moulin's game requires Adam and Eve to begin by simultaneously naming a probability. The player who wins this auction, by bidding the highest probability p, then gets to propose an outcome[13] s. If the other player assents to this proposal, a referee then organizes a lottery that generates the final result. The lottery yields s with probability p, and the status quo d with probability $1-p$. If the other player refuses, she gets to make a counter-proposal which, if accepted, leads to the same rigmarole as before (with the *same* probability p). If the counter-proposal is refused, d occurs for certain.

Moulin's game has been given in detail to make clear how "unnatural" its rules are. Who organized this game and why? Where did the referee come from? And, most importantly, *what constrains the players to obey the rules?*

As regards bargaining in Rawls' "original position" (Part I, section 2), and in most other contexts also, one wants to *minimize* on the constraints hypothesized as restricting the players' freedom of action. One can argue that the Nash demand game epitomizes this minimizing aim. Rational players choose the course of action that maximizes expected utility given their beliefs. If they are not shackled by informational constraints, they will know exactly what to believe and what to prefer. What course of action will they then choose in a bargaining situation? If they have the freedom to make commitments (*binding* promises), it is to be anticipated that they will do so at the earliest possible moment, leaving themselves no room for future maneuver, so as to confront their opponent with a take-it-or-leave-it problem. Thus all action will be telescoped into the first instant as modeled in the Nash demand game.

But I do not think this is a good argument. It is true that one does not want to invent artificial constraints on the players' freedom of action. But nor does one wish to invent artificial freedoms of opportunity. Much of Part I is concerned with this point and so comment here will be restricted to the remark that to *announce* a commitment is not to *make* a commitment. As Schelling [1960] demonstrates, commitments need an *enforcement mechanism* to make them stick. But, to hypothesize an enforcement mechanism without explanation, is to run afoul of the minimizing principle.

One must therefore proceed *without* primitive commitment assumptions. This does not mean that threats, for example, are no longer to be regarded as relevant: only that, where threats are made, an endogenous explanation must be offered of why they are credible. You may *say* that you will do something if a future contingency occurs, but why should I believe you if what you threaten will not be optimal *after* the contingency is realized? Notice that *time* has now entered the picture. No longer can it be argued that all of the action can be seen as telescoped into a single instant.

To illustrate these points, imagine the Nash demand game, without uncertainty, modified so that Adam makes his demand *first*. Eve is then left with a take-it-or-leave-it ultimatum. If Adam's ultimatum leaves her more than her status quo payoff, rationality says that she must accept.[14] To express this more grandiosely, the only equilibrium in the one-player *subgame* which follows such an ultimatum from Adam requires Eve to make the maximum demand compatible with that already made by Adam. Adam, knowing that Eve will behave rationally in these subgames, therefore optimizes by minimizing the amount that his demand leaves for Eve. Thus Adam scoops the pot.[15]

Of course, with the Nash bargaining solution demands n_A and n_E, Adam does *not* scoop the pot. Nevertheless, the pair (n_A, n_E) is a ***Nash equilibrium*** for the ultimatum game (along with many others). If Eve is *always* going to use the strategy of demanding n_E whatever Adam does, then Adam cannot gain by demanding more than n_A. The point here is that the definition of a Nash equilibrium is too crude to capture the fact that a threat by Eve to play this way is just not credible. When *time* is relevant, the Nash equilibrium idea needs to be *refined.*

The purpose of analyzing the ultimatum game is to introduce Selten's [1975] notion of a ***subgame-perfect equilibrium.*** The definition is very simple. The equilibrium strategies must constitute a Nash equilibrium, not only in the whole game, but *in every subgame as well.*[16]

5. Rubinstein's bargaining game. (Woe to Ariel..., Isaiah, 29,1).

Why would Eve not make a counter-bid in the ultimatum game after rejecting Adam's original bid? Only because the rules forbid it. Rubinstein [1982] (see also Stahl [1972]) formulated a game without such an artificial constraint. Adam and Eve alternate in making proposals until a proposal is accepted. Hitherto it has been the players' attitudes to ***risk*** which has been crucial, but now their attitudes to ***time*** must also be considered. As Cross [1969] remarks, "If it did not matter *when* the parties agreed, it would not matter *whether* they agreed at all."

In the discussion that follows, the assumption[17] will be that a deal s reached at time t is worth $\delta_A^s \varphi_{A(s)}$ to Adam and $\delta_E^t \varphi_E(s)$ to Eve. The numbers δ_A and δ_E are therefore ***discount factors.*** They are assumed to lie between 0 and 1. The corresponding discount ***rates,*** ρ_A and ρ_E are defined by $\delta_A = e^{-\rho_A}$ and $\delta_E = e^{-\rho_E}$. A time interval of $\tau > 0$ is assumed to separate consecutive proposals, the first of which is made by Adam at time 0. A utility of zero is assigned by both players to the event that agreement is ***never*** reached.

Rubinstein's result is that such games have a *unique* subgame-perfect equilibrium. (See chapters 3, 5 and 6 of Binmore/Dasgupta [1987]: particularly p.109 for a quick proof for the current special case.) Rubinstein's model is certainly a natural one, but it still incorporates unexplained constraints. Why does Adam go first? Why do the players politely wait their turn? Who chose the time interval of τ between successive proposals?

Such difficulties evaporate when the limiting case $\tau \to 0+$ is considered. This is the interesting case for a minimizing program because a player will want his or her next bid on the table as soon as possible after a rejection has been registered. But, in the limiting case, the problem of who goes first disappears since the first mover advantage becomes vanishingly small.

The limiting case of Rubinstein's game *minimizes* on the fictions which need to be promulgated to explain the behavior of those playing the bargaining game. In particular, the rules of the game leave players with no obvious motive for cheating. One does not therefore need to hypothesize exogenously determined penalties for infringements of the rules. Nor is it necessary to hypothesize an exogenous mechanism for sustaining commitments. It is therefore a striking vindication of Nash's insight that the unique equilibrium outcome in the limiting case as $\tau \to 0+$ is a weighted *Nash bargaining solution.* To see this is not hard, given that there is a unique subgame-perfect equilibrium.

With $\tau > 0$, let $P = (a, b)$ be the equilibrium outcome in the Rubinstein game with Adam as first–mover. If Adam's opening proposal is refused, then a subgame is entered in which Eve is the first–mover. Let $Q = (c, d)$ be the equilibrium outcome in the Rubinstein game if Eve were first–mover. Since the subgame after Adam's opening proposal is refused begins at time $t = \tau$, the equilibrium outcome of the subgame is necessarily $(\delta_A^\tau c, \delta_E^\tau d)$. It follows that, in equilibrium, Eve must accept any proposal $(x, y) \in X$ made by Adam for which $y > \delta_E^\tau d$, because the latter is what she will get from refusing. Similarly she will refuse if $y < \delta_E^\tau d$. The equilibrium proposal by Adam must therefore be $y = \delta_E^\tau d$. Similar considerations apply when it is Eve who proposes. Thus,

$$\left.\begin{aligned} b &= \delta_E^\tau d \\ c &= \delta_A^\tau a \end{aligned}\right\} \qquad (5.1)$$

Raise both sides of the first of these equations to the power $\beta_E = 1/\rho_E$. Do the same for the second equation using the power $\beta_A = 1/\rho_A$. Multiplying the equations then yields the result that

$$a^{\beta_A} b^{\beta_E} = c^{\beta_A} d^{\beta_E}$$

and hence the points $P = (a, b)$ and $Q = (c, d)$ both lie on the same contour $x^{\beta_A} y^{\beta_E} = k$, where k is a constant. The situation is illustrated in fig.3A (in which $X_\tau = \{(\delta_A^\tau x, \delta_E^\tau y) : (x, y) \in X\}$).

Fig.3B shows the limiting case as $\tau \to 0+$. The points P and Q converge to a common limit n at which a contour $x^{\beta_A} y^{\beta_E} = k$ touches the frontier of X. It follows that n is the weighted Nash bargaining solution for $(X, 0)$ with bargaining powers $\beta = 1/\rho_A$ and $\beta_E = 1/\rho_E$.

Figures 3A and 3B here

Aside from providing a minimalist defense of the Nash bargaining solution, the model also yields some input on the interpretation of the "bargaining powers" in the asymmetric case. These do *not* quantify differing bargaining "skills". One cannot do better than be a rational optimizer, and this is assumed of both players. The asymmetries, in the Rubinstein game, arise from their differing attitudes to the unproductive passage of time. The more impatient player has less bargaining power. For simplicity, in the examples which follow, only the symmetric case $\rho_A = \rho_E$ is considered.

6. Impasse and exit. The study of specific bargaining models does more than provide a rationale for the use of the Nash bargaining solution. It offers clues on *how* it should be used.

The following example will be relevant to the discussion of bargaining in the original position which appears in part V of this program of papers. But the story will be told here in the more familiar circumstances of a wage negotiation.

The first point of interpretation is that bargaining often concerns *flows* rather than *stocks* as has been taken for granted so far. But, since a flow of u utils per period is equivalent to a utility stock of $u + u\delta + u\delta^2 + \ldots = u/(1-\delta)$, where a fixed discount factor of δ $(0 < \delta < 1)$ is assumed, one can easily transfer the analysis to the language of flows. In particular, there is no difficulty in regarding d as a pair of utility flows and X as a set of pairs of utility flows when this is convenient.

In a wage negotiation, the set of feasible deals will be derived from the wage and profit flows that can result from a successful agreement. But what of the disagreement point d? At least two candidates present themselves. The first candidate will be called the *exit point* e. It is the pair of utility flows which will result if the negotiations are broken off irrevocably as a consequence of one player abandoning the attempt to reach an agreement and taking up his or her best *outside option* instead. This will leave the other player with no choice but to do the same thing. The other candidate will be called the *impasse point* i. This is the pair of utility flows that the players get while they are negotiating but have yet to reach an agreement[18]—i.e. during a *strike*. In section 5, $i = e = 0$.

It is traditional in the labor economics literature to locate the disagreement point d at the exit point e and then to use the symmetric Nash bargaining solution. But, without implausible commitment assumptions, there is little to be said in favor of this practice.

If the Rubinstein bargaining model is modified (as detailed in Binmore/Shaked/ Sutton [1988]) to allow for the possibility of exit, the disagreement point d *remains* at the impasse point i as in section 5. The equilibrium outcome n, in the limiting case $\tau \to 0+$, is the weighted Nash bargaining solution for the problem (Y, i), where $Y = \{x : x \in X \text{ and } x \geq e\}$, as illustrated in fig.4A for the case $\rho_A = \rho_E$. Observe that the exit point e serves only as a *constraint* on the range of validity for the Nash bargaining solution.

Figures 4A and 4B here

7. Bargaining without trust. What enforces a deal once it has been struck? For breaches of commercial contracts, a doubtful remedy may be sought via the legal system. For the social contract, corresponding appeals are usually made to "natural law". Part I, section 7 explains why I am unwilling to follow this line. In brief, obligations and duties are part of what a social contract theory should *explain.* To embed them in an *a priori* system of "natural law" is to beg a central question.

This section is therefore concerned with studying deals which are honored because it is *in the interests* of the parties to the agreement to honor them—i.e. the act of honoring the agreement is *in equilibrium.* Such a requirement has, for example, implications for the shape of the set X. As explained in part I, section 8(ii), the assumption that X is convex becomes a *substantive* assumption for which the justification in the current paper is *only* that it simplifies the analysis. The shaded region in fig.4B is the set of subgame-perfect equilibrium outcomes[19] for the Rubinstein bargaining model (in the case $\rho_A = \rho_E$ and $\tau \to 0+$) for the *non-convex* feasible set X illustrated. One of the problems to be faced without convexity, is therefore a loss of *uniqueness* in this model. In persisting with my assumptions on X, I am therefore guilty of a substantial evasion.

The shape of X is one problem. But there are others. Consider, in particular, the following archetypal example of bargaining without trust. Two criminals agree to exchange a quantity of heroin for a sum of money. The buyer is to end up with the seller's heroin and the seller with the buyer's money. But how is this transition to be engineered if each is free to walk away at any time, without penalty, carrying off whatever is currently in his or her possession?

Obviously, there is no point in the buyer handing over the agreed price and waiting for the goods. Somehow the criminals have to arrange a flow between them, so that the money and the drug change hands *gradually*.[20] Moreover, the rate at which heroin is exchanged for money at any time must be regulated so that it never be in the interests of either criminal to call a halt to the exchange in order to *renegotiate* the terms of trade. At each time, therefore, the deal which would be negotiated, given the *current* distribution of goods, must lead to the same final outcome as the original deal. Otherwise the agreed final outcome would not have been included in the feasible set in the first place, there being no viable way of implementing it.

The situation is illustrated in fig.5A. The point e_0 is the original exit point –i.e. the no-trade point. Trade is seen as the movement of the exit point along a curve joining e_0 and the final outcome n. The disagreement point d is *not* identified with the no-trade point e_0. Impasse consists of hanging around on some street corner dickering, at constant risk of attracting the attention of the ever-vigilant forces of law and order.[21] The point e represents a transitional exit point, some money and some heroin having changed hands. Notice that, with an impasse point i that remains fixed during the transaction, *any* curve joining n and e_0 *along which utility steadily increases* for both Adam and Eve will suffice as a trading path.

Figures 5A and 5B here

Fig.5B shows a somewhat different situation. Here the utilities are to be interpreted as flows (rather than stocks as in the heroin-trading example). It illustrates a species of game played by the workers and the management in operating a firm. The game is assumed to have many equilibria.[22] Each point in X corresponds to the pair of utility flows that results from the choice of a particular equilibrum. Currently the firm is operating the Pareto-inefficient equilibrium corresponding to i_0. Meanwhile management and workers bargain about how to share the surplus that would result if agreement could be reached on a move to a more efficient equilibrium. Thus i_0 is the impasse point. A very inferior exit point is indicated at e_0. With the disagreement point at i_0, the Nash bargaining solution is located at n. How does the firm get from i_0 to n without the existence of trust between management and workers?

Such a move will involve the gradual surrender of entrenched privileges and practices on both sides. Once surrended, these will not necessarily be recoverable in their old form. If the move to n is broken off before it is completed, perhaps by one side or the other insisting on a renegotiation, the firm will lapse into a *new* equilibrium. The point i in fig.5B corresponds to such an equilibrium. If there is a renegotiation demand, it will be such a point i that serves as the *new* impasse point. The agreement can therefore be seen as a deal to move the *impasse point* i along a curve joining i_0 to n. (One should not ask whether such a curve exists: the set X should be *defined* as the set of endpoints of such curves).

The requirement that the final outcome originally agreed remains the agreed final outcome after a renegotiation with a new impasse point i constrains the curve that can be used to link i_0 and n very sharply. Only a *straight line*[23] will suffice, as illustrated in fig.5B.

It is this last story which is most relevant to the social contract discussion of part V. What I want to insist upon here is that it is pointless to be using the Nash bargaining solution, or any other bargaining solution, without a very clear idea of the circumstances in which it is to be applied. In particular, it would definitely be *wrong*, given the management-worker story that goes with fig.5B, to be placing the disagreement point at e_0—i.e. at the analog of a Hobbesian state-of-nature—unless commitment assumptions are to be made. This is a point which has been insisted upon by Buchanan [1975,1976].

8. Continuous exchange. Rubinstein's bargaining model treats the case when *no* commitment is possible as the limiting case of a situation when tiny commitments *are* possible (lasting for a time period of τ). In treating the problem of trustless cooperation in terms of movement along a continuous path, as in the previous section, one implicitly relies on a similar modeling device. Each agent is assumed to be trusting in respect of tiny amounts, and then the limiting case is considered. However, the game-theoretic considerations are not so straightforward as in the Rubinstein model.

In the heroin-trading example of the previous section, suppose that the buyer has $100, each dollar of which is worth only one cent to him if not spent on heroin. Each unit of heroin purchased by the buyer for one dollar, at the agreed rate of exchange, is worth only one cent to the seller if not sold to the buyer. Equating utils with dollars and aping the alternating move structure of the Rubinstein model, one obtains the game of exchange illustrated in fig.6. At each decision node, the player whose turn it is to move can choose "across" or "down". To choose "across" is to make a gift to the other player which is worth one cent to the donor, but one dollar to the recipient. To choose "down" is to cheat on the arrangement by exiting with what one currently has.

This game has only one subgame-perfect equilibrium. This requires both players *always* to cheat. No trade would then take place. To see this, consider what is optimal in the subgame that arises if the rightmost (final) node is reached. The seller must then choose between 100.01 and 100, and so cheats by choosing the former. In the subgame that arises if the penultimate node is reached, the buyer will predict this behavior by the seller. He must therefore choose between 99.01 and 99, and so cheats by choosing the former. Since the same "backwards induction" argument works at every decision node, the result of a subgame-perfection analysis is that both players plan always to cheat.

Figure 6 here

Does this mean that exchange is not possible without trust? To draw such a conclusion would be to put more weight on the mathematical model proposed above than it can bear. The real world is imperfect in many different ways. The model of fig.6 takes account of the imperfection that money is not infinitely divisible. But real people are even more imperfect than real money. In particular, they are not infinitely discriminating. What is one cent more or less to anybody? Following Radner [1986], one may seek to capture such an imperfection by looking at *ϵ-equilibria* in which players are satisfied to be within ϵ of the optimal payoff. Provided that the relevant trading unit (one cent in fig.6) is chosen smaller than ϵ, it will then be an ϵ-equilibrium to honor the deal. One can then proceed to idealize the situation by allowing $\epsilon \to 0+$, hence obtaining a continuous[24] model of trade. Applying this technique to the ultimatum game at the end of section 4, one finds many ϵ-equilibria. Adam's ultimatum must take account of the fact that Eve may not accept a proposal which assigns her ϵ or less. Similarly, there are many ϵ-equilibria in Rubinstein's model of section 5. In the limit as $\epsilon \to 0+$, all these equilibria converge on the unique subgame-perfect equilibrium. But this is *not* true in the trading game of this section. In particular, honoring the deal is a limit of ϵ-equilibria. For this paper, nothing more is necessary. Cheating is *also* the limit of ϵ-equilibria, but it is not a limiting equilibrium on which the agents would wish to coordinate. This may leave the reader unsatisfied. Is cheating *really* viable for rational players? Perhaps towards the end of the game: but surely not right from the beginning? This is essentially the same question that arises in the repeated *Prisoners' Dilemma.* Indeed, the trading game of this section was introduced by Rosenthal [1981] as a simplified version of the repeated *Prisoners' Dilemma.* It is generally agreed that game theory has yet to provide an adequate resolution of the difficulties raised by such examples. My own view [Binmore,1987] is that a resolution depends on taking what it means to be an imperfect human being very much more seriously than is customary. Playing around with ϵ-equilibrium sheds light on only one small corner of the mystery. Fortunately for these papers, this is the corner in which we are sitting.

9. Information. This is a large subject on which only very little will be said. One large evasion throughout these papers will be that only games of *complete* information are ever considered. Players always know the rules of the game and the tastes and beliefs of their opponents. However, information is not necessarily *perfect*—i.e. players will not necessarily know the full history of moves in the game so far. Nor are *chance moves* excluded.

These last points are essential in modeling the Rawlsian *original position.* Players 1 and 2 are hypothesized as bargaining behind a *veil-of-ignorance* which conceals their real-life roles as Adam or Eve. They treat these roles as being determined by a *chance move* which attaches a probability of 1/2 to each of the two possible role-assignments. The intuition is that deals reached in the original position will then be "fair". (See Part I for an extended discussion). Information is *imperfect* in the original position because it is players 1 and 2 who are making decisions and *their* tastes and beliefs, while making the decisions, are assumed to be common knowledge. A better theory would leave room for incomplete information in the original position. But progress in this direction must await a more solidly founded theory of bargaining with incomplete information than is currently available.

These clarifying remarks are necessary in order to clear the air for a brief discussion of a rather different informational question. In section 2, the weighted Nash bargaining solution was characterized by three axioms. But there is a further "implicit axiom". It is that the bargaining solution should depend *only* on the mathematical entity (X, d). Information which is abstracted away in this idealization of the bargaining problem is to be suppressed altogether. It is tempting to follow Harsanyi [1977,p118] in brushing aside the relevance of other parameters as "strategically irrelevant". Those who find this airy defense adequate are invited to skip forward to section 11.

Interest in these matters has been revived by Roemer [1988], but it is earlier work which will be significant here. I am particularly concerned to clarify the relation between the Nash bargaining solution and a Walrasian allocation. The latter is what results from trading at market-clearing prices in an exchange economy. It is therefore a concept which lies close to the heart of conservative thinkers.

10. Walrasian solution. Fig.7A shows an Edgeworth box representing a two-person trading situation with two commodities, wheat and fish. The no-trade or endowment point is located at e. There is a unique Walrasian outcome at w. The slope of the line joining e and w is $-p$, where p is the clearing price of fish in terms of wheat. The symmetry of the configuration implies that the allocation corresponding to the symmetric Nash bargaining solution lies at n (provided that the disagreement point, which need not be at e, is symmetrically located).

Figures 7A and 7B here

In section 2, a "bargaining problem" was narrowly defined as a pair (X, d). But this conventional definition is an uncomfortable straitjacket which is abandoned in what follows. This makes it possible to regard the Walrasian outcome in the Edgeworth box of fig.7A as a type of *bargaining solution.* As such, it differs from the Nash bargaining solution in taking account of a *wider* informational base. It depends, not only on utility information, but on all the information in the Edgeworth box.[25] In chapter 10 of Binmore/Dasgupta [1987] (also Gevers [1986]), it is argued that a Walrasian outcome is, in fact, the appropriate *extension* of the symmetric Nash bargaining solution to such an *expanded* informational base.

Recall, from section 2, the proposition characterizing the weighted Nash bargaining solution. What follows is an expanded and amended list of the characterizing properties:

1. Strong individual rationality
2. Independence of utility *and commodity* calibration[26]
3. Independence of irrelevant alternatives[27]
4. Pareto-efficiency
5. Symmetry

An analog of the proposition of section 2 can now be stated. The analog concerns functions f which map each Edgeworth box into an allocation in the box. If such a function satisfies properties 1–5, then its values are Walrasian.

Bargaining solutions are under discussion in this paper because it is necessary to resolve the bargaining problem faced by those in the original position. Which bargaining solution is chosen will depend on what is hypothesized about the circumstances of those in the original position. Part I took the view that the usefulness of the device of the original position depends on *minimizing* the fictions to be promulgated. Hence the concern in the current paper with whether the rules of a given bargaining procedure would or would not be honored if there were no penalties for cheating.

Given this attitude, a critic might quote the characterization of the Walrasian correspondence given above, or some other such result, and challenge the use of the Nash bargaining solution in the original position. Why should those in the original position be denied information about the actual physical goods that equity may require to be redistributed? Such critics may include, not only those for whom defending the market mechanism is a knee-jerk reflex, but also those who have sought to characterize equity in terms of *envy-free* allocations (e.g. Varian [1974,1985], Thomson and Varian [1984]).

An allocation is envy-free if no individual would wish to swop the bundle of resources allocated to him or her with that allocated to someone else. In a simple exchange economy, like that considered above, an envy-free allocation may be obtained as follows. First confiscate all the original endowments and then redistribute these so that each agent gets an equal amount of each commodity. This creates a *new* endowment point e as illustrated in fig.7B. The Walrasian allocation relative to this new endowment point e is then envy–free. Thus w in fig.7B is one of the envy-free allocations.

Such criteria evade the vexed question of how inter-personal comparisons of utility should be made (section 9 and part I, section 6). But the evasion entails a heavy cost, as examples like the following make clear. A benefactor presents two bottles of vermouth and twenty bottles of gin to Adam and Eve on condition that they agree on how to split the donation between them. Adam mixes Martinis only in the ratio of 1 part of vermouth to 100 parts of gin. Eve mixes Martinis only in the ratio of 1 part of vermouth to 10 parts of gin. Neither has any use for gin or vermouth beyond mixing such Martinis. An equal split of the donation gives 1 bottle of Vermouth and 10 bottles of gin to each. Any trade makes Eve worse off and hence no trade will take place. That is to say, the equal-split endowment is already the Walrasian allocation. But one can hardly say that it is "fair" in any meaningful sense. It is true that Adam does not envy Eve's allocation and that Eve does not envy Adam's, but Eve is able to enjoy 10 times as many Martinis as Adam.

But it is not adequate simply to criticize potential critics. It remains necessary to offer some justification for neglecting all but utility information. A consideration of what is involved in implementing the Walrasian solution w in fig.7A will assist in clarifying this point. The simplest bargaining model that does the trick is an analog of the Nash demand game. But, instead of utility demands, the players announce the maximum amount they are willing to trade of the good they wish to sell, and the minimum rate (price) at which they will exchange this good for the good they wish to buy. Trade is then assumed to take place at the maximum volume consistent with the announced constraints. In fig.8A, the set C represents the set of allocations consistent with Crusoe's announcement, and F the set consistent with Friday's announcement. The set of allocations consistent with both is $C \cap F$. The allocation representing maximal trade in this set is v. A Pareto-efficient, Nash equilibrium in ths game is necessarily Walrasian. The sets C^* and F^* in fig.8A represent equilibrium strategies. (Chapter 10 of Binmore/Dasgupta [1987]).

Figures 8A and 8B

The next step is to indicate what strategies in this primitive trading game look like in terms of *utilities*. Fig.8B illustrates the set U of utility pairs corresponding to the set C of allocations in fig.8A. A strategy in the trading game amounts to *demanding* a utility pair in such a set U. All that needs to be observed is that, if Crusoe picks such a demand set, Friday would definitely *not* wish to be restricted to picking a demand set of a similar odd shape. In the circumstances of the original position, to *constrain* Friday in this way would be to conflict with the *minimizing* principle on unexplained fictions. The moral of the story is intended to be that, unless unexplained constraints are imposed on those in the original position, information other than utility information will be *irrelevant*, just as Harsanyi was quoted as saying towards the end of section 8.

11. Utilitarian solution. This is usually treated as a social choice rule, rather than a bargaining solution, as in Thomson [1981]. That is to say, given a feasible set X, the utilitarian solution is normally identified with the point x in X at which $x_A + x_E$ is maximized. The treatment of this section takes as fundamental, not just a feasible set X, but a *pair* (X, d). The set X has the same interpretation as in the case of the Nash bargaining solution (section 2), with the gloss that it will be assumed *strictly* convex (to assure that the utilitarian solution gets defined *uniquely*). However, d will no longer be interpreted as a disagreement *point*, but as a disagreement *direction*.[28] Only the interesting case $d > 0$ will be considered.

Fig.9A illustrates the weighted utilitarian solution u for the problem (X, d) corresponding to the weights $\omega_A > 0$ and $\omega_E > 0$. This is the value of $x = (x_A, x_E)$ at which the "weighted sum"

$$\sigma = \frac{\omega_A x_A}{d_A} + \frac{\omega_E x_E}{d_E}$$

is maximized subject to the requirement that $x \in X$. Notice that only the ratio $\omega_A d_E / \omega_E d_A$ is significant. Notice also the close resemblance between the form of the definition and that given for the weighted Nash bargaining solution in section 2.

Figures 9A and 9B here

This resemblance is no accident. Recall the proposition of section 2 that characterizes the Nash bargaining solution. If property 1 is replaced by Pareto-efficiency and (X, d) is reinterpreted as in the current section, then *the proposition characterizes the weighted utilitarian solution.* Not only this, a simple adaptation of the Rubinstein bargaining game of section 5 leads to a utilitarian outcome. Suppose that a deal s reached at time t is not evaluated, as in section 5, as equivalent to the utility pair $(\delta_A^t \varphi_A(s), \delta_E^t \varphi_E(s))$, but as equivalent to $(\varphi_A(s) - tc_A, \varphi_E(s) - tc_E)$. That is to say, the players do not *discount* the unproductive passage of time but pay a *fixed-cost* per unit of time lost. Recall that τ denotes the time interval between successive proposals. In the limiting case when $\tau \to 0+$, the unique subgame-perfect equilibrium outcome converges on the unweighted utilitarian solution with $c = (c_A, c_E)$ as the disagreement *direction*.[30] It is such models that motivate the interpretation of the line marked "disagreement path" in fig.9A. If agreement were indefinitely delayed but rational behavior were always anticipated in the future, the expected deal would recede along the disagreement path towards an "infinitely distant impasse".

An intuition for the result can be obtained by supposing the existence of discount factors $\delta_A < 1$ and $\delta_E < 1$ along with the fixed costs. Adam's *impasse payoff* is then

$$-c\tau_A\{1 + \delta_A^\tau + \delta_A^{2\tau} + \ldots\} = -c_A^\tau/(1 - \delta_A^\tau) \to -c_A/\rho_A \text{ as } \tau \to 0+ .$$

Following the principle of section 6, one should therefore employ the weighted Nash bargaining solution with bargaining powers $\beta_A = 1/\rho_A$ and $\beta_E = 1/\rho_E$, locating the disagreement point at $d = (-c_A/\rho_A, -c_E/\rho_E)$. Now let $\rho_A \to 0+$ and $\rho_E \to 0+$ in such a way that $\rho_A/\rho_E \to \omega_E/\omega_A$. The solution then converges to the weighted utilitarian solution for the problem (X, c) with weights ω_A and ω_E as illustrated in fig.9B.

In part I, section 2, Harsanyi's [1977] defense of utilitarianism as "fair" was described. In the version given of his argument, the Nash bargaining solution was used to resolve the bargaining problem faced by those in the original position. But, if the aim is to achieve a utilitarian outcome, why fuss about with the original position in the first place? Why not allow Adam and Eve to bargain *without concealing* *their identities* and resolve their bargaining problem using the utilitarian solution directly? I believe that to ask this question is to misunderstand the whole nature of the enterprise. Bargaining solutions are not like ready-made suits. You cannot go into a shop and buy whichever takes your fancy. A bargaining solution needs to be carefully tailored to the problem it supposedly resolves. The Nash program is a tool for checking how well a bargaining solution fits a particular problem. What it tells us about the utilitarian solution is that it is a hopeless misfit for Adam and Eve's problem. Other issues aside, they begin with a disagreement *point* (corresponding to the state-of-nature): not a disagreement *direction.*

12. Inter-personal comparison. This subject has already been introduced in the martini example of section 9. Who gets how many martinis is clearly relevant to what is or is not "fair". But suppose Adam likes Martinis very much more than Eve. Perhaps Eve then "deserves" more Martinis to compensate for this disadvantage. Or perhaps Adam should get all the Martinis to maximize total "welfare". It is clear that some theory of inter-personal comparison of utility is necessary to tackle such issues conveniently. The orthodox methodology, on which my approach is founded, employs the notion of "extended sympathy" as described in part I, sections 4–6. The current section reviews some of the primitive ideas.

In part I, section 4, two states h (hell) and H (heaven) are introduced to anchor the utility scales. Adam's and Eve's utility functions are then normalized so that $\varphi_A(h) = \varphi_E(h) = 0$ and $\varphi_A(H) = \varphi_E(H) = 1$. Can we now say that both players will have gained equally if they find themselves both in heaven rather than in hell? Obviously not, since the calibration of utility scales using the two benchmark states is entirely *arbitrary.*

Sometimes, such considerations are allowed to confuse bargaining discussions. For example, the status quo state s_0 (section 2) in a Nash bargaining context is said to provide a "natural" hell-point - i.e. a location for the zero point on the players' utility scales. Often, this is expressed by saying that the description of the bargaining problem determines a "natural" comparison of utility *levels*.

The same trick may be played with the utilitarian solution. Rather than using a disagreement *point* to provide a "natural" comparison of *levels*, the disagreement *direction* may be used to provide a "natural" comparison of utility *units*.

If one were provided with both a disagreement point *and* a disagreement direction then one would have a "natural" comparison of *both* utility levels and utility units. Thus a *full* inter-personal comparison of utilities would be achieved. Fig.10A illustrates the bargaining solution that would necessarily result from maintaining the interpretations of disagreement point and disagreement direction used in this paper. The result is called a *proportional bargaining solution* in the literature (Raiffa [1953], Isbel [1960], Kalai [1977], Myerson [1977], Roth [1979]). The axiomatization which perhaps best captures the fact that utility scales are to be deemed as fully tied down is that of Peters [1986].

Figures 10A and 10B here

Returning to the case of the Nash bargaining solution, consider the following piece of window-dressing. Having located the hell-point at the status quo, it is "natural" to locate the heaven-point at the state t implemented by the Nash bargaining solution. This is, after all, the best that rational bargainers can achieve. Hence so the argument goes, the Nash bargaining solution is "fair" because each bargainer gains one util.

Although I agree that this naive argument has no merit, I do think that it makes good sense to seek to *endogenize* the issue of inter-personal comparison, along the lines explained in Part I. I also believe that a *context-dependent* notion of "fairness" should not be too quickly rejected. Real-world usages certainly have this property.[31]

The naive argument is offered in protest against the use of the word "natural" to disguise statements which are not even clearly meaningful. In the Von Neumann and Morgenstern theory, the natural (i.e. primitive) entities are preference relations. The utility functions are simply mathematical constructs and their properties should be *deduced* from properties of the ***primitives*** of the theory - i.e. the preference relations. The fact that it is a fallacy to argue that an agent prefers x to y *because* $\varphi(x) > \varphi(y)$ is widely understood. When it is argued that a deal is "fair" *because* of inter-personal comparison reasons, I believe we should suspect that a similar fallacy is in train.

A critic might respond that this is all very well for Von Neumann and Morgenstern utility functions, which are derived from attitudes to risk. These are essential for an approach based on the original position, because of the hypothetical chance move which determines who occupies which role. But suppose that the utilities are hypothesized as representing "levels of desert" or "intensities of need". My response to this is that one simply cannot *announce* that such an interpretation is to be made. A theory, analogous to that of Von Neumann and Morganstern, needs to be provided which explains the principle by means of which these fuzzy notions are to be measured on a scale like temperature. How else is one to know what value judgements may or may not be hidden in the woodwork?

13. Kalai/Smorodinsky solution. (And Ehud said, I have a message..., *Judges*, 3, 20)

Given a Nash bargaining problem (X, d), begin by calculating the most to which each player can aspire. This is the largest payoff consistent with the other player getting at least his or her disagreement payoff. The resulting payoff pair i will be called the *ideal point.* Next construct a "disagreement path" by joining d and i with a straight line. The Kalai/Smorodinsky solution k is located where this line crosses the Pareto-frontier of X, as shown in fig.10B.

Similar bargaining solutions (with different ideal points) appear in Part VI. However, in this section I want to argue against using the Kalai/Smorodinsky solution as an *alternative* to the Nash bargaining solution. Gauthier [1986], for example, seeks to justify this practice by using an idiosyncratic version of Zeuthen's principle. However, I do not see why the concession stage of his process is not essentially a Nash demand game and the claim stage therefore irrelevant.

What of Kalai and Smorodinsky's [1975] defense? They have no bargaining models, only axioms. Nash's *independence of irrelevant alternatives* is replaced by a *monotonicity* axiom (usually now called "individual monotonicity"). Suppose that the disagreement and ideal points are left fixed, but the feasible sets change so that, for each of Eve's possible utility demands, there will be more left in the new situation for Adam if he concedes the demand than there would have been in the old situation. The monotonicity axiom says that Adam should get more in the new situation than in the old. The grounds are that his bargaining position has improved. But what basis is there for this claim? It is certainly *false* as a general proposition. The Nash and Rubinstein bargaining models provide *counter-examples.* Models do exist in which the claim is valid. The model of Moulin [1984] mentioned in section 4 provides an example. But such models are remote from those that come immediately to mind in a bargaining context.

14. Morals by agreement. This section itemizes the cornerstones of the theory that Gauthier [1986] espouses in has book "Morals by Agreement". His aims coincide with mine but we differ radically on several vital points in seeking to achieve this aim, and hence in the conclusions at which we arrive. The current section attempts to clarify the points of difference by considering briefly each of the five "conceptions" that Gauthier regards as basic to his approach [1986, pp13–17]. These are taken in reverse order from that in which he introduces them.

1. The Archimedean Point. The first notion is that of an "Archimedean point" [Rawls, 1971, p584] from which perspective moral issues can be viewed impartially. That is to say, some analog of the Rawlsian original position is required. I agree wholeheartedly with Gauthier that it is not satisfactory to examine the conclusions that an "ideal observer" would reach under such conditions of impartiality. A rational bargaining analysis would seem necessary if the results are to be convincing.

2. The state of nature. The second notion concerns the *status quo* from which bargaining begins in Gauthier's story. I agree that to neglect consideration of this point would be a serious defect. Wolff [1977] has been forceful in criticizing Rawls on this issue, and a similar criticism can be directed at Harsanyi [1977]. Neither Rawls nor Harsanyi see any necessity to refer to an original "state of nature" and hence individual rationality constraints do not arise in their treatments.

However, I am not comfortable with the manner in which Gauthier envisages the "initial bargaining position" being selected [1986, p190]. Although it is easy to see why he wishes to introduce the criteria he proposes, the arguments offered in defense of the criteria seem to me to have an *ad hoc* flavor. I am not even comfortable with the idea that the initial bargaining position should be seen as an object that can be selected. To my mind, the manner in which the Archimedean point is specified should render the initial bargaining position inevitable.

There is also a minor technical point that arises on this topic. In part I, it was explained how the use of Nash's threat game will lead to a Hobbesian state–of–nature being used as *status quo* under appropriate circumstances. However, I do not believe that the Nash threat game is an appropriate tool for use in the original position because of its reliance on commitment assumptions (see item 3 below). But Gauthier [1986, p200] is not entitled to dismiss the approach as he does in the following quote:

> ... fortunately we need not face the problem of determining maximally effective threat strategies ... They play a purely

> hypothetical role in the Nash–Harsanyi analysis, since Ann and Adam do not actually choose them ... But if Ann and Adam would not choose these strategies, then they cannot credibly threaten with them. Maximally effective threat strategies prove to be idle.

If such arguments concerning the use of credible threats were valid, game theorists would have to abandon concepts like that of subgame–perfect equilibrium. It is true that rational bargainers never actually employ the threat strategies that they commit themselves to use if agreement should fail to be reached. This is because rational bargainers always reach agreement in this story. But *what* they agree on depends on what would happen if they were to fail to agree. That is to say, it depends on the threats the players have made in advance. As always, the reason the players stay on the equilibrium path is because of what they anticipate happening if they were to deviate. Since rational players never deviate, what would happen if they were to deviate always remains hypothetical. But it is far from irrelevant to their actual behavior.

3. Commitment. The third notion is that of "constrained maximization". It is true that "better" societies could be designed if it were possible for individuals to make unbreakable commitments that constrained their future behavior. Moreover, there is no shortage of situations in which individuals, either acting alone or as part of a group, would find it in their interests to make such commitments. What I believe to be indefensible is the proposition that such commitments *can* be made in the absence of a suitable enforcement mechanism. Without such a mechanism, any hypothetical agreements reached in an analog of the original position will have to be *self-policing*. That is to say, only *equilibria* will be available as the objects of a rational agreement. One may seek to evade this admittedly unpalatable conclusion by postulating "natural laws" that serve to police hypothetical agreements that are not self-policing. But this does not advance matters very far, since the policing of the natural laws themselves is still left unexplained.

This issue seems to me to be central for a social contract theory. However, the preceding paragraph is all that will be offered on the subject here. More is said in part I of the sequence of papers of which this is part II. In brief, I think Hume [1739, p280] was right in saying:

> What theory of morals can ever serve any useful purpose, unless it can show that all the duties it recommends are also in the true interst of each individual?

Gauthier [1986, p1] makes this his opening sentence as a proposition requiring immediate denial.

4. Bargaining. The fourth notion is Gauthier's concept of "minimax relative concession" from which he deduces a bargaining solution equivalent to that of Kalai and Smorodinsky [1977] in two-person situations. The bulk of this paper has been devoted to expounding the orthodoxy he thereby outrages. In so far as it was ever the case that this orthodoxy was the "Nash–Harsanyi–Zeuthen" theory that he offers as the principle alternative his account of rational bargaining [1986, p133], this is no longer true. It is correct that it remains orthodox to defend the Nash Bargaining Solution, but the basis for this defense is now very much more securely grounded in non-cooperative game theory. In particular, it is no longer considered adequate to appeal to *ad hoc* criteria like "Zeuthen's Principle".

Gauthier [1986, p133] envisages rational bargaining as a two-stage process in which each bargainer makes a claim followed by a concession. Suppose there are just two bargainers whose utilities for the initial bargaining position are u_1^* and u_2^*. Each is then seen as making a claim, $u_1^\#$ and $u_2^\#$. The rational claims, so Gauthier asserts, are those that maximize $u_1^\#$ and $u_2^\#$ subject to the requirement that

$(u_1^*, u_2^{\#})$ and $(u_1^{\#}, u_2^*)$ correspond to feasible agreements. The bargainers then make concessions according to an unspecified process. This leads to agreement on a utility pair (u_1, u_2). Gauthier defines the "relative magnitude of a player's concession" as $c_i = (u_i^{\#} - u_i)/(u_i^{\#} - u_i^*)$. He emphasizes the scale–invariance of this quantity. He next insists that rational bargainers will select (u_1, u_2) so as to ensure that the larger of c_1 and c_2 is minimized. This is the "principle of minimax relative concession".

His defense of this last requirement is very brief. He asserts [1986, p143] that the condition

> . . . expresses the equal rationality of the bargainers. Since each person, as a utlity–maximizer, seeks to minimize his concession, then no one can expect any other rational person to be willing to make a concession if he would not be willing to make a similar concession.

In this quote, the word"concession" is used in two different senses. At its first appearance, no more tham Gauthier's algebraic defintion seems to be intended. It is, of course, tautologous that a maximizer of u_i is a minimizer of c_i. At its later appearances, it has acquired a more complex meaning that needs more careful examination.

It is certainly attractive to entertain the proposition that no rational person would expect another rational person to take an action that he himself would not take under *identical* circumstances. Having granted this, it is tempting to extend the proposition by requiring that no rational person would expect another rational person to take an action that he himself would not take under *equivalent* circumstances. The question then arises: when are the circumstances under which rational players operate equivalent? Gauthier's answer seems to be: when the sets of concessions from which they choose are the same, other structure being irrelevant. It may be that such a proposition could be defended if the definition of the word "concession" were tailored to the situation. But Gauthier offers no defense at all for the use of his simple algebraic definition in this context. It is, indeed, difficult to see what form such a defense would take without specific reference being made to the concession *strategies* available to the players.

One route might be to suppose that, after making their claims, each player simultaneously announces a take–it–or–leave–it concession. But with such a model, it is difficult to see why the final outcome would be as sensitive to the initial claims as Gauthier requires. Indeed, since the initial claims in Gauthier's story are entirely stereotyped, it is not clear why he accords them so much prominence. Rational players have no need to interchange information that is common knowledge between them already.

5. Morally–free zone. Gauthier [1986, p13] argues that a perfectly competitive market constitutes a morally–free zone. The view seems to be based on the observation that a Walrasian equilibrium is Pareto–efficient.

I do not want to comment on this point in the specific context of Gauuthier's argument, where it seems to me to be something of a side–show. But it does seem worthwhile to take the opportunity of forestalling some misunderstandings that such a view may provoke.

Both Gauthier and I are concerned with "morals by agreement". Although the arguments may be couched in abstract terms, the question being asked is essentially a practical one: how can we get from where we are to somewhere better *by mutual consent*. This is to take a much narrower view then is commonplace of what appeals to morality can realistically achieve, with my view being substantially more pessimistic than Gauthier's. My approach is to model the circumstances under which societies operate as a game, and to see morality as a system of rules for selecting and sustaining *equilibria* in that game. I have my doubts about the extent to which a free market in which large numbers of agents persistently end up with inadequate access to basic necessities can be said to be "in equilibrium". Why should the have–nots feel constrained to honor the property rights of the haves if they have the power to do something about it? This, however, is to quibble. What is important is that a system of morals that operates ***by mutual consent*** cannot shift society from a Pareto–efficient equilibrium because at least one person would be made worse off and hence would not cooperate.[32] The most that one can ask of a practical system of morals in a rational society is that it provide guidance while society is shifting from an inefficient equilibrium to an efficient one. I am aware that such a view leads to conclusions that sound more than a little Panglossian. But it seems to me that this is inevitable if morality is to be treated as one of the "arts of the possible".

Footnotes

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

2. Although Rawls would deny this.

3. Alekhine is reported as saying, "Position, what does position matter? It is my will that counts." To model him as *homo economicus* would therefore be as fruitless as to do the same for the chimpanzee.

4. The set S should be assumed to be closed under the formation of lotteries when deals based on these are feasible so that attitudes to risk are properly taken account of.

5. The formal definitions are: (1) *Closed*: X contains its boundary points; (2) *Bounded above*: there is a vector b such that $x \in X$ implies $x \leq b$; (3) *Convex*: if X contains x and y, then it contains the line segment joining them; (4) *Comprehensive*: if X contains x and y, then $x \leq z \leq y$ implies $z \in X$.

6. Arrow's theorem is not immediately applicable. For one thing, the condition of *unrestricted domain* is not appropriate here. Thus, for a fixed d, it is possible to construct a collective preference relation on $\{x : x > d\}$. With properties 1,2 and 3 of the proposition, the indifference curves have the form $(x_A - d_A)^{\beta_A}(x_E - d_E)^{\beta_E}$ = constant. But then Adam and Eve would be prepared separately to take risks that they would reject collectively. This seems unacceptable.

7. For those familiar with Maskin's [1985] *monotonicity*, a preference relation can be defined on $\{x : x > d\}$ as follows. Make Adam indifferent between d and any $x \notin X$ (i.e. any non-feasible payoff pair). If x and y are both in X, make Adam strictly prefer x *to* y if and only if $x_A > y_A$. Proceed similarly for Eve. With these preferences, the independence of irrelevant alternatives follows from monotonicity.

8. See Maskin [1985] and Moore/Repullo [1988]. Abreu/Sen [1987] suggests that matters are perhaps more complex.

9. Section 2 does not say that utility scales should not be used: only that they should not be abused.

10. If the choice of incompatible demands led to the payoff pair d, the result would be the same except that $(Y, 0)$ becomes (Y, d). Technical matters are discussed in chapters 4 and 8 of Binmore/Dasgupta [1987], especially pp.159–167. See also Carlsson [1987].

11. Sometimes, it is argued that Nash's axioms should be construed as *selection criteria* for this situation. This is a coherent view, but not one that is very helpful in the current context. Why these selection criteria and not others?

12. This is a lot of supposing. A more careful discussion is really necessary here.

13. Which may itself be a lottery.

14. One is not entitled to argue that she would not *want* to accept a "derisory" offer. She is *hypothesized* to be a maximizer of expected utility. In real life, she might easily turn down a derisory monetary offer in favor of nothing without being irrational. But this would imply that her utility for the latter option was greater in the circumstances of the decision.

15. If Adam is able to make arbitrarily low offers, Eve will actually get nothing at all in *equilibrium*. It cannot be optimal to offer Eve even a tiny amount if she is going to accept half as much. Of course, in real-life, goods come in discrete units.

16. Whether or not, the subgame will actually be reached in equilibrium. Binmore [1987] warns against an *uncritical* espousal of the idea.

17. Fishburn and Rubinstein [1982] give axioms relating to attitudes to time (rather than risk, a la Von Neumann and Morgenstern) which justify such functional forms.

18. The impasse point can be neatly characterized as follows. If the bargainers agreed upon the impasse point, they would not care *when* they agreed upon it. (Binmore/Rubinstein/Wolinsky [1986]).

19. See chapter 5 of Binmore/Dasgupta [1987].

20. Or to be more accurate, so that *control* over the money and drug changes hands continuously. One sometimes sees exchanges in the street in which the traders grasp both of the objects being exchanged and gradually relinquish the firmness of their hold. Of course, in real-life, there is always the threat of physical violence in the background.

21. Suppose Adam attaches utility b_A to being picked up by the police at time 0 while engaged in a drug transaction. If the probability of this happening in any time interval of length τ is $\pi\tau$, then $i_A = b_A\pi\tau\{1 + (1-\pi\tau)\delta_A^{\tau} + (1-\pi\tau)^2\delta_A^{2\tau} + \ldots\} = b_A\pi\tau/\{1-(1-\pi\tau)\delta_A^{\tau}\} \rightarrow b_A\pi/(\rho_A + \pi)$ as $\tau \rightarrow 0+$.

22. Only stationary equilibria are considered.

23. This is the same straight line which appears when Zeuthen's principle is used, but I do not see that much significance can be attached to this.

24. Why not model directly in continuous time? In brief, one cannot do this directly, as Shubik [1983,p60] proposes, because mathematical coherence then requires a well-ordering structure on each play of the game. One may turn to the theory of differential games within which all is mathematically sound. But one does not escape the necessity of considering imperfections. These simply find their way into the specification of the state equations where they are hard to evaluate.

25. Which is to be understood as encompassing not only a region in commodity space but also Von Neumann and Morgenstern utility functions on this region.

26. Independence of commodity calibration means that the solution should not depend on the units in which commodities are measured.

27. See chapter 10 of Binmore/Dasgupta [1987].

28. When a vector d is used in physics to specify a direction, the formal understanding is that it is a representative of the equivalence class $D = \{\lambda d + \varepsilon : \lambda > 0$

and $e \in \mathbb{R}^2\}$. Thus, in section 10, (X, d) is to be understood as shorthand for what is *formally* (X, D).

29. Strong individual rationality is not meaningful when there is no disagreement *point*.

30. See chapter 5 of Binmore/Dasgupta [1987]. Rubinstein [1982] uses the fixed-cost case to illustrate the existence of multiple equilibria, but his X is then not strictly convex.

31. See Binmore/Shaked/Sutton [1988] for some experimental evidence.

32. One should not be naive about what is subsumed under the notion of an equilibrium. It will not usually be in equilibrium for the wicked to go unfettered about their business if their behavior is sufficiently offensive to those about them provided that the latter have the power to interfere. Nor will there usually be reason to curb the activities of those who have a proclivity for altruism. However, *everybody*, both wicked and saintly, will be cooperating in sustaining the equilibrium because it is in what they see as their interests to do so.

References

D. Abreu and A. Sen [1987], "Virtual implementation in Nash equilibrium," mimeo, Harvard University.

K. Binmore [1984], "Game theory and the social contract I," ICERD discussion paper 84/108, London School of Economics.

K. Binmore [1987], "Modeling Rational Players I," *J. Econ. and Phil.*

K. Binmore [1988], "Game theory and the social contract II," ICERD discussion paper 88/170, London School of Economics.

K. Binmore and P. Dasgupta [1987], *The economics of bargaining,* Basil Blackwell, Oxford.

K. Binmore, A. Rubinstein and A. Wolinsky [1986], "The Nash bargaining solution in economic modeling, "*RAND. J. of Econ.* 17, 176–187.

K. Binmore, A. Shaked and J. Sutton [1988a], "An outside option experiment," mimeo, London School of Economics.

K. Binmore, A. Shaked and J. Sutton [1988b], "Do people exploit their bargaining power?" forthcoming.

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.

H. Carlsson [1987], "A bargaining model where parties make errors," discussion paper, University of Lund, Sweden.

J. Cross [1969], *The economics of bargaining,* Basic Books, New York.

P. Dasgupta, P. Hammond and E. Maskin [1979], "The implementation of social choice rules: some general results in incentive compatibility, *Review Econ. Studies*, 46 185–216.

P. Fishburn and A. Rubinstein [1982], "Time preference," *Int. Econ. Review*, 23, 677–694.

D. Gauthier [1986], *Morals by agreement,* Clarendon Press, Oxford.

L. Gevers [1986], "Walrasian social choice: some simple axiomatic approaches", *Social Choice and Public Decision Making, Essays in Honor of Kenneth J. Arrow.* Vol. 1, ed. W. Heller, R. Starr, D. Starre H, C.U.P., Cambridge.

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

J. Harsanyi [1955], "Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility," *J. Political Econ.* 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.P., Cambridge.

J. Isbell [1960], "A modification of Harsanyi's bargaining model," *Bull. Amer. Math. Soc.* 66, 70–73.

E. Kalai [1977], "Proportional solutions to bargaining situations: interpersonal utility comparisons," *Econometrica* 45, 1623–1630.

E. Kalai and M. Smorodensky [1975], "Other solutions to Nash's bargaining problem," *Econometrica* 45, 1623–1630.

E. Maskin [1985], "The theory of implementation in Nash equilibrium: a survey," in *Social goals and social organization* ed. L. Hurwicz, D. Schmeidler, H. Sonnenschein, Cambridge U.P., Cambridge.

J. Moore and P. Repullo [1988], "Nash implementation: a full characterization", *ST/ICERD working paper TE/88/175*, London School of Economics.

H. Moulin [1984], "Implementing the Kalai-Smorodinsky bargaining solution," *J. Econ. Th.* 33, 32–45.

R. Myerson [1977], "Two-person bargaining and comparable utility," *Econometrica* 45, 1631–1637.

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

J. Nash [1951], "Non-cooperative games," Annals of Math. 54, 286–295.

J. Nash [1953], "Two person cooperative games," Econometrica 21, 129–40.

H. Peters [1986], *Bargaining game theory,* Proefschritt Universitat Nijmegen, Netherlands.

R. Radner [1986], "Can bounded rationality resolve the Prisoners' Dilemma?" *Essays in Honour of Gerard Debreu*, eds. A. Mass-Collel and W. Hildenbrand, North-Holland, Amsterdam.

H. Raiffa [1953], "Arbitration schemes for generalized two-person games," *Contributions to the Theory of Games II*, ed. Kuhn and Tucker, Princeton U.P., Princeton.

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

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

J. Rawls [1972], *A theory of justice*, Oxford U.P., Oxford.

J. Roemer [1988], "Axiomatic bargaining theory in economic environments", *J. Econ. Th.*, forthcoming.

R. Rosenthal [1981], "Games of perfect information, predatory pricing and the chain-store paradox," *J. Econ. Th.* 25, 92–100.

A. Roth [1979], *Axiomatic models of bargaining,* Lecture notes in economics and mathematical systems 170, Springer-Verlag, Berlin.

A. Rubinstein [1982], "Perfect equilibrium in a bargaining model," *Econometrica* 50, 97–100.

T. Schelling [1960], *The strategy of conflict*, Harvard U.P., Cambridge, Mass.

R. Selten [1975], "Re-examination of the perfectness concept for equilibrium in extensive-form games", *Int. J. Game. Th.* 4, 25–55.

M. Shubik [1983], *Game theory in the social sciences: concepts and solutions*, M.I.T. Press, Cambridge, Mass.

I. Stahl [1972], *Bargaining theory*, Economics Research Institute, Stockholm.

R. Sugden [1986], *The economics of rights, cooperation and welfare*, Basil Blackwell, Oxford.

W. Thomson [1981], "Nash's bargaining solution and utilitarian choice rules," *Econometrica* 49, 535–538.

W. Thomson and H. Varian [1984], "Theories of justice based on symmetry."

H. Varian [1974], "Equity, envy and efficiency," *J. Econ. Th.* 9, 63–91.

H. Varian [1985], "Dworkin on equality of resources," *Economics and Philosophy* 1, 110–125.

J. Von Neumann and O. Morgenstern [1944], *The theory of games and economic behavior*, Princeton U.P., Princeton.

R. Wolff [1977], *Understanding Rawls*, Princeton University Press, Princeton.

F. Zeuthen [1930], *Problems of monopoly and economic warfare*, Routledge and Kegan Paul, London.

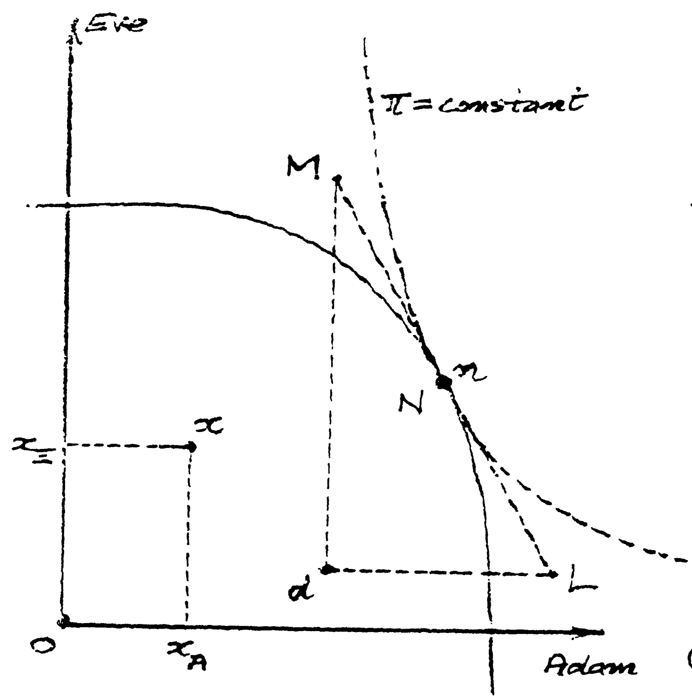

figure 1A

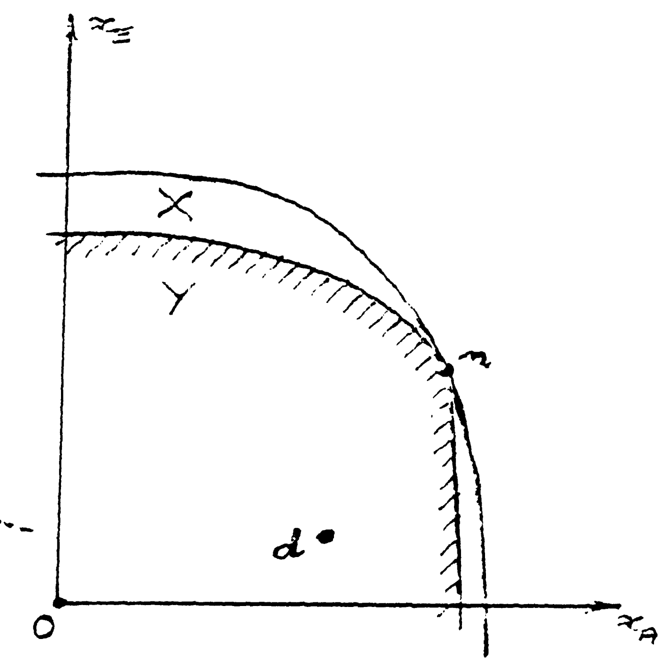

figure 1B

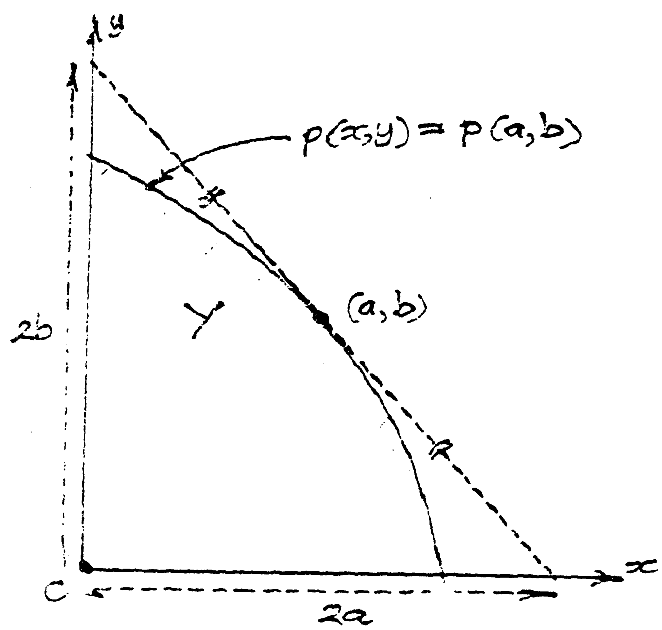

figure 2A

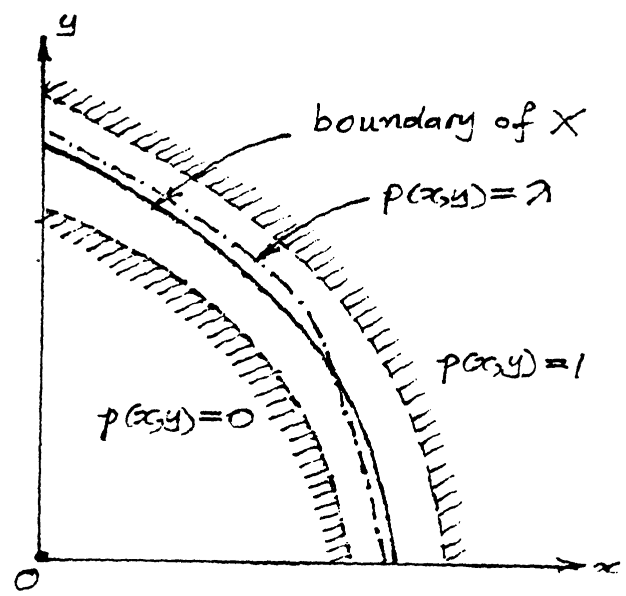

figure 2B

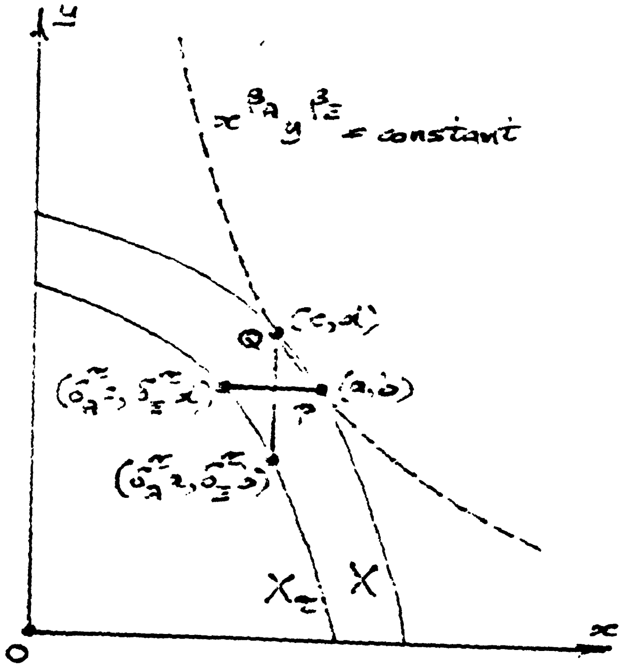

figure 3A

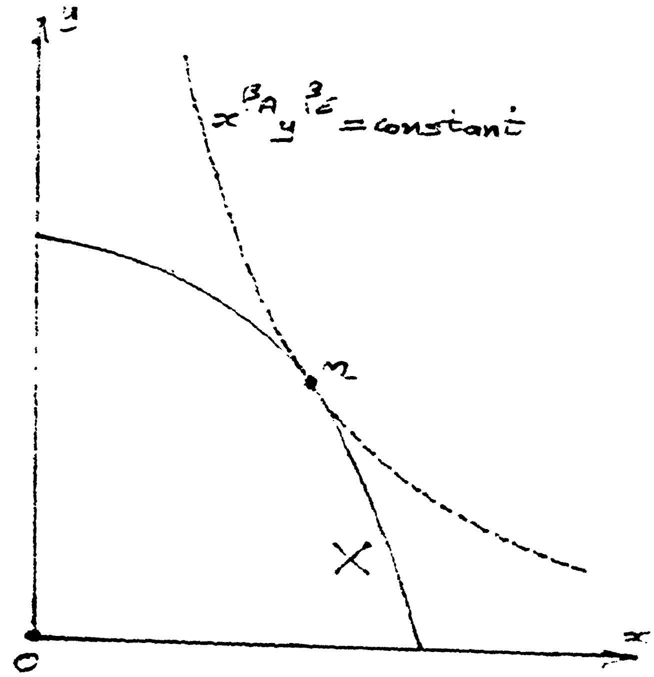

figure 3B

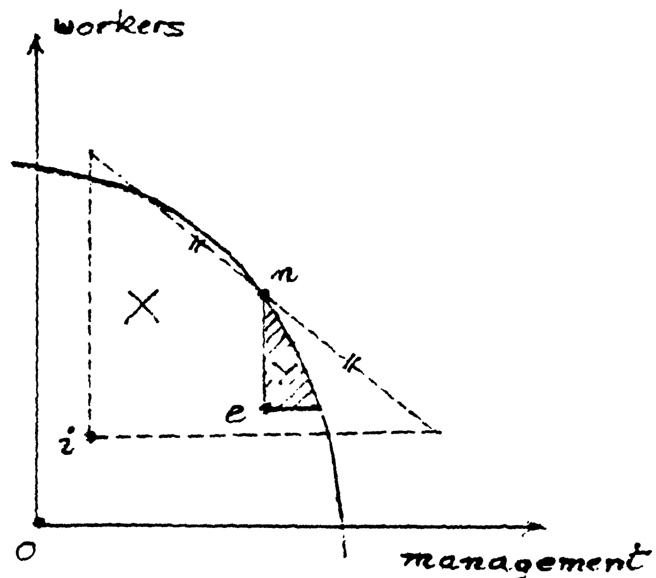

figure 4A

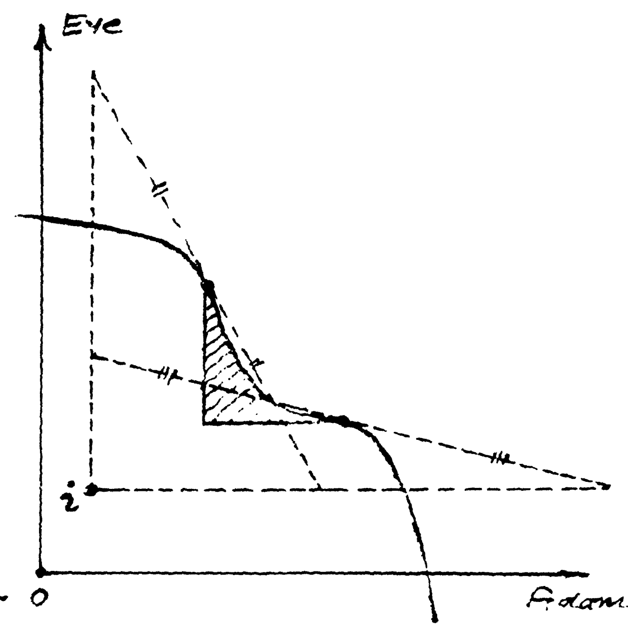

figure 4B

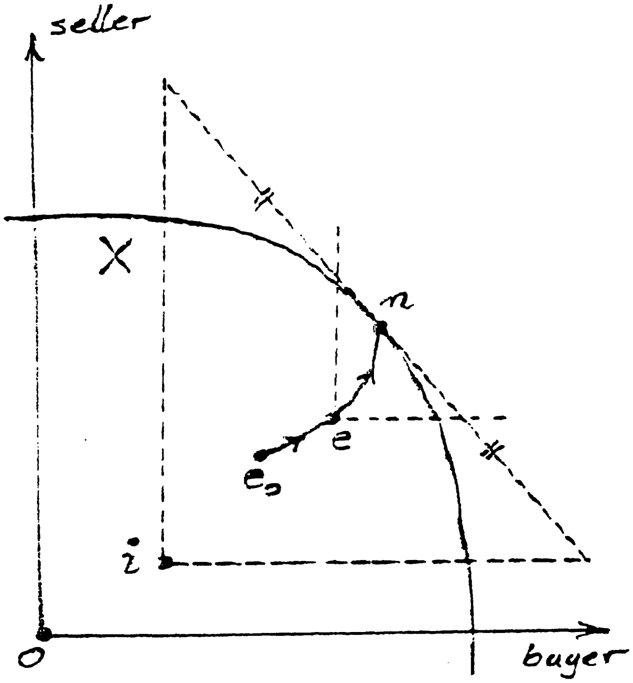

figure 5A

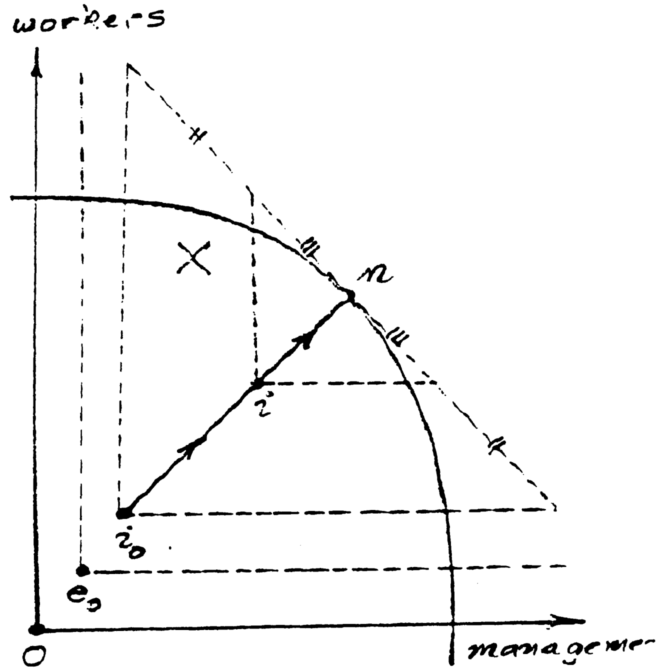

figure 5B

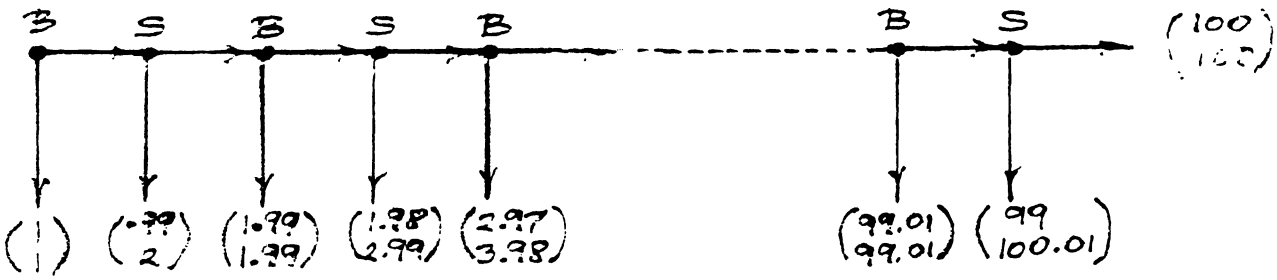

figure 6

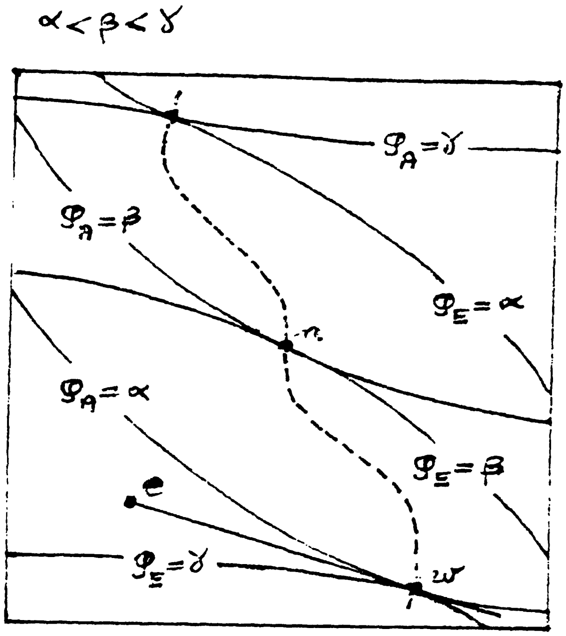

figure 7A

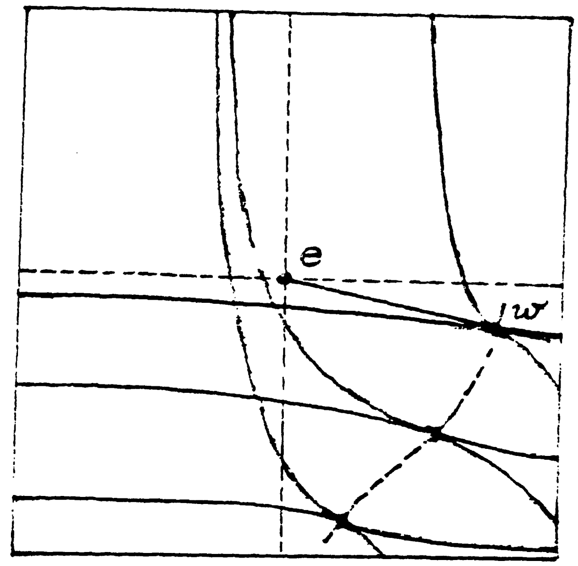

figure 7B

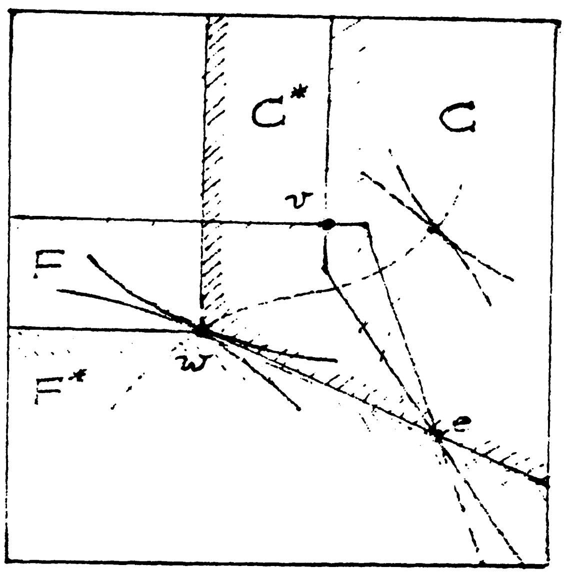

figure 8A

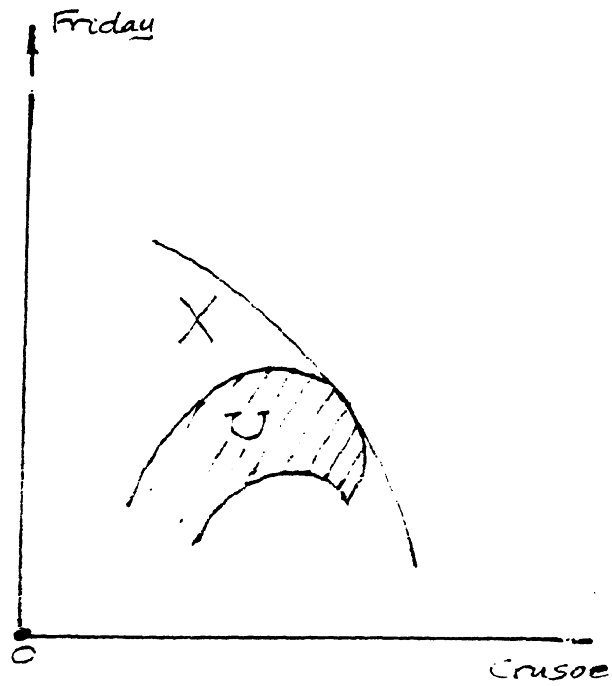

figure 8B

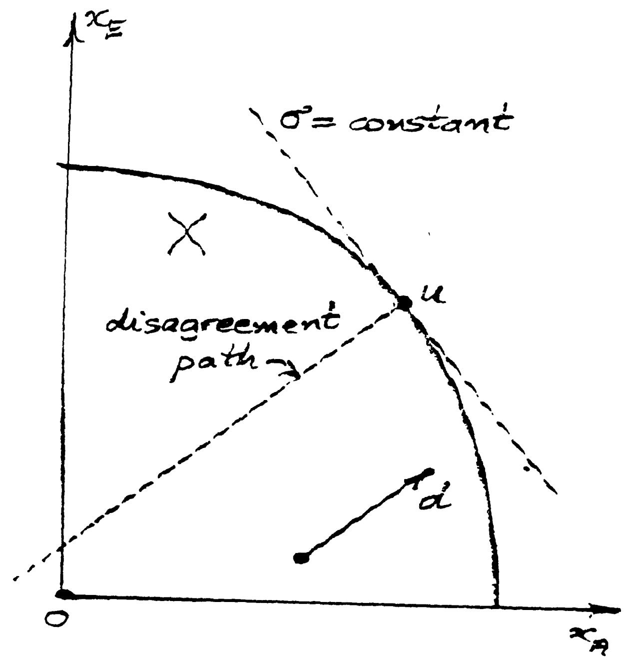

figure 9A

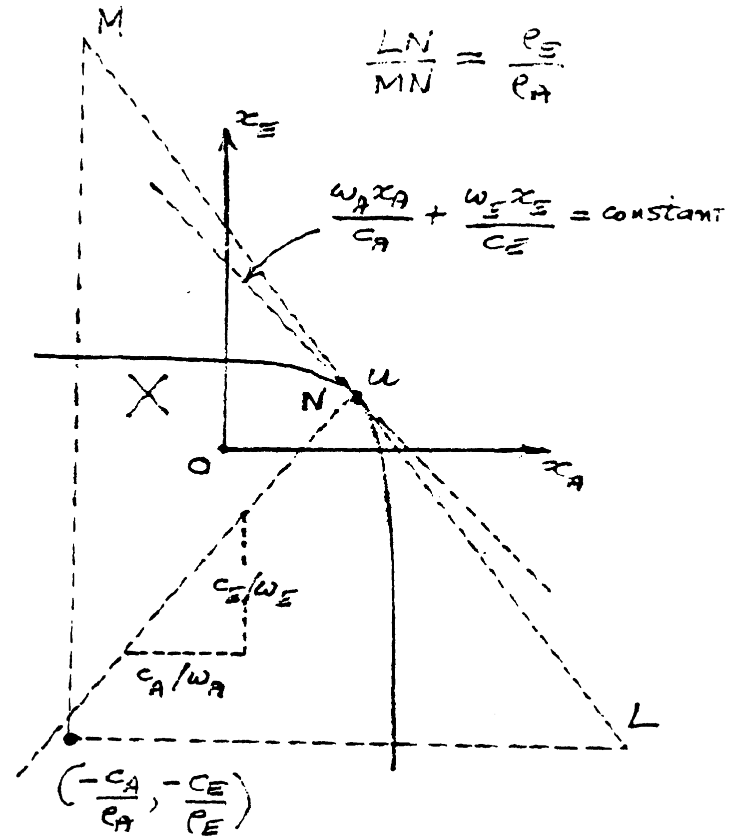

figure 9B

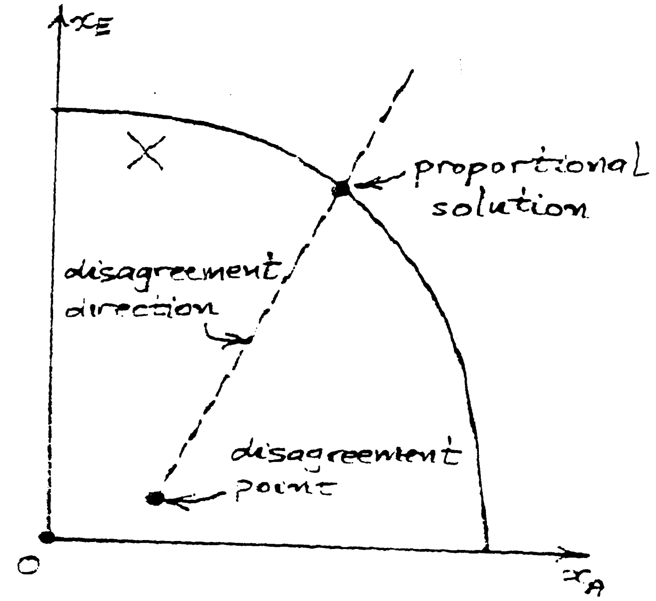

figure 10A

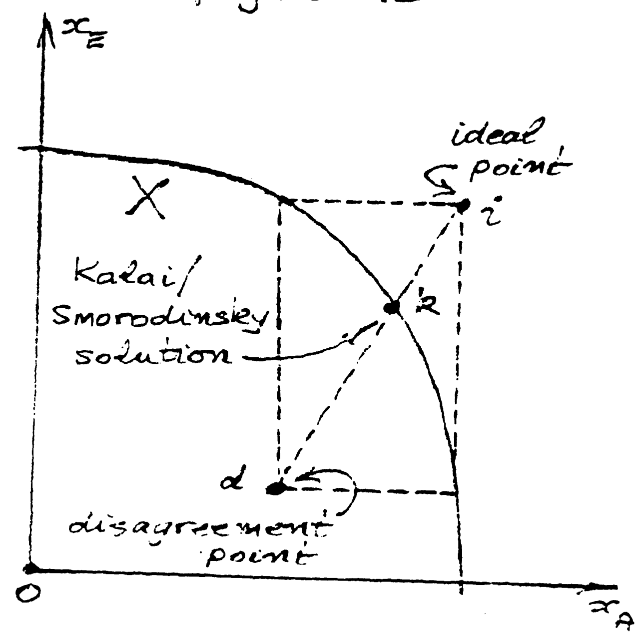

figure 10B

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.

DATE DUE
