Words: 1777

Chapter 4. Other forms of Kantian optimization

4.1 A continuum of Kantian equilibria

Multiplicative and additive Kantian optimization each employ a method of 'universalizing one's action.' We can generalize this as follows. Consider the function

$$\varphi^{\times}(E,r) = rE \ . \tag{4.1}$$

defined on the domain $\Re_+^2$. Let a game be defined by payoff functions $\{V^i\}$ on strategy profiles $(E^1,...,E^n)$. We can define a multiplicative Kantian equilibrium as a profile $(E^1,...,E^n)$ such that:

$$(\forall i) \quad V^i(\varphi^{\times}(E^1,r),...,\varphi^{\times}(E^n,r)) \text{ is maximized at } r=1 \ . \tag{4.2}$$

Similarly, define:

$$\varphi^{+}(E,r) = E + r - 1 \ . \tag{4.3}$$

Then an additive Kantian equilibrium is a profile $(E^1,...,E^n)$ such that:

$$(\forall i) \quad V^i(\varphi^{+}(E^1,r),...,\varphi^{+}(E^n,r)) \text{ is maximized at } r=1 \ . \tag{4.4}$$

(In this case, *r* can take on negative values.)

More generally:

Definition 4.1 A function $\varphi(E,r):\Re_+ \times \Re \to \Re_+$ such that $\varphi(E,1) \equiv 1$ is a *Kantian variation.*

Consider the convex economic environments with production $e=(u^1,...,u^n,G)$ that we have been working with: denote the domain of such environments by $\mathfrak{E}$. The environment *e* becomes an *economy* if we append to it an allocation rule, which is a set of functions $\{X^i:\Re_+^n \to \Re_+, i=1,...,n\}$ such that:

$$\forall (E^1,...,E^n) \quad \sum_{i=1}^{n} X^i(E^1,...,E^n) = G(E^S) \ .$$

Alternatively we may define an allocation rule as a set of output share functions θ^i, where $X^i(E^1,...,E^n) = \theta^i(E^1,...,E^n,G)G(E^S)$. We have studied two allocation rules, $X^{\Pr}$ and X^{ED} . Given a pair (e,X) we have an economy with respect to which we can

define a game whose strategies are the effort levels of its members. What we have shown is that the (positive) Kantian equilibria of economies (e,X^{Pr}) with respect to the Kantian variation $\varphi^{\times}$ are Pareto efficient for all $e \in \mathfrak{E}$, and the Kantian equilibria of economies (e,X^{ED}) with respect to the Kantian variation φ^{+} are Pareto efficient for all $e \in \mathfrak{E}$. This motivates the definition:

Definition 4.2 A pair (X,φ) consisting of an allocation rule and a Kantian variation will be called an *efficient Kantian pair* if the Kantian equilibria with respect to the variation φ of economies (e,X) for all $e \in \mathfrak{E}$ are Pareto efficient.

The question naturally arises: are there any efficient Kantian pairs other than these two, where *e* belongs to the class of convex economic environments $\mathfrak{E}$? The answer is there is a whole continuum of such pairs, which span a set of which the two polar members are $(X^{\mathrm{Pr}},\varphi^{\times})$ and (X^{ED},φ^{+}). There are, in addition, some others that I discuss in the next section.

Define the allocation rule, for any $\beta \in [0,\infty)$:

$$X_{\beta}^{i}(E^{1},...,E^{n}) = \frac{E^{i}+\beta}{E^{S}+n\beta}G(E^{S}) \tag{4.5}$$

and the Kantian variation:

$$\varphi_{\beta}(E,r) = rE + (r-1)\beta . \tag{4.6}$$

Notice that as $\beta \to \infty, X_{\beta}(E^{1},...,E^{n}) \to \frac{G(E^{S})}{n}$. Let us therefore define $\varphi_{\infty}(E,r) = \varphi^{+}(E,r)$. Notice that for $\beta = 0$, we have the proportional allocation rule, and the multiplicative Kantian variation. We have:

Proposition 4.1 *For all* $0 < \beta \leq \infty$, $(X_{\beta},\varphi_{\beta})$ *is an efficient Kantian pair on* $\mathfrak{E}$.

Proof:

1. An effort vector $\mathbf{E} = (E^{1},...,E^{n})$ is a Kantian equilibrium in the economy $(u^{1},...,u^{n},G,X_{\beta})$ with respect to the Kantian variation φ_{β} if and only if:

$$(\forall i=1,...,n)(\frac{d}{dr}\Big|_{r=1} u^i(\frac{E^i+\beta}{E^S+n\beta}G(rE^S+n(r-1)\beta),rE^i+(r-1)\beta)=0)\,. \qquad (4.7)$$

(To verify (4.7), compute that the fraction $\frac{E^i+\beta}{E^S+n\beta}$ is invariant with respect to application of the function φ_β to all the effort levels.) Now (4.7) expands to:

$$u_1^i\cdot\frac{E^i+\beta}{E^S+n\beta}G'(E^S)(E^S+n\beta)+u_2^i\cdot(E^i+\beta)=0\ , \qquad (4.8)$$

which reduces to:

$$u_1^iG'(E^S)+u_2^i=0\ , \qquad (4.9)$$

which uses that fact that $E^i+\beta>0$ because $\beta>0$. This proves the claim for $\beta\in(0,\infty)$.

2. The case $\beta=\infty$ -- which is K^+ equilibrium – has been shown in Proposition 3.3.

■

We see from the last part of step 1 of the proof why we do not require the restriction to *positive* Kantian equilibria, for $\beta>0$, that is required in Proposition 3.2.

Prop. 4.1 demonstrates the existence of a continuum of efficient Kantian pairs spanning economies from the fishing ($\beta=0$) to the hunting ($\beta=\infty$) economies. What do these allocation rules X_β look like? Let's write:

$$\frac{E^i+\beta}{E^S+n\beta}=\lambda\frac{E^i}{E^S}+\frac{(1-\lambda)}{n}\ , \qquad (4.10)$$

and compute that its solution in λ is:

$$\lambda^*=\frac{E^S}{E^S+n\beta},\quad (1-\lambda^*)=\frac{n\beta}{E^S+n\beta}\ . \qquad (4.11)$$

λ^* is independent of *i,* so (4.11) implies that:

$$(\forall i=1,...,n)\quad x^i=\lambda^*\frac{E^i}{E^S}G(E^S)+(1-\lambda^*)\frac{G(E^S)}{n}. \qquad (4.12)$$

We can describe these allocations as follows. They are Pareto efficient, and a fraction λ^* of the product is divided in proportion to effort, while the rest is equally divided among the participants. The value of λ^* is endogenous – it depends upon what the Kantian equilibrium is. We do know, of course, that as β travels from 0 to infinity, λ^* travels from one to zero. But we cannot specify *a priori* a particular convex combination λ we wish to implement and immediately choose the right β. That is to say, the mapping from λ to β is complicated, depending on E^S. ($\beta = \frac{E^S(1-\lambda)}{\lambda}$.)_ Indeed, these rules have appeared quite often in axiomatic resource-allocation analysis: see Moulin (1987), Ju et al (2007), Chambers et al (2015), and Thomson (2015).

It is interesting that the allocation rules that can be implemented efficiently on the domain $\mathfrak{E}$ are essentially only the 'convex combinations' of the two classical cooperative rules – equal-division of the product and division of product in proportion to labor expended[1]. Classically, Karl Marx associated these rules with 'communism' and 'socialism'. Communism, for Marx, was allocation of the product in proportion to need, which in the case of equal needs is equal division, while socialism was allocation in proportion to labor expended. Marx lived before Pareto described his concept of efficiency, and so could not have been expected to demand efficiency as well as a desideratum of cooperative society. Kantian optimization has a natural pedigree in history of thought of cooperative societies – enriched by Pareto.

There is a geometric representation of the family of Kantian variations φ_β : see Figure 4.1, which illustrates the case $n = 2$. At an allocation of effort (E^1, E^2), both players contemplate expanding the allocation along the ray through the origin labeled $K^\times$ if they are applying the multiplicative rule; if they are applying the additive rule, they contemplate changing the allocation along the 45^0 line labeled K^+. As β travels from 0 to infinity, they contemplate changing the allocation along a dotted line, as illustrated, lying in the cone generated by K^+ and $K^\times$. On the other hand, if the players are Nash optimizers, then the first one contemplates changing the effort allocation along

[1] Section 4.2 qualifies this statement.

the horizontal line labeled N^1 and the second contemplates changing it along the vertical line labeled N^2 . One might ask: Are there more complex variations then these linear ones that would enable the efficient implementation of other allocation rules than the ones described? The answer is no. (Theorem 3, Roemer (2015).)

I do not think there is any practical application of Kantian optimization for the rules given by $0 < \beta < \infty$. The ways of thinking required by the variations φ_β for this open interval are too complicated. This chapter is of purely theoretical interest: its main conclusion is that Kantian thinking is naturally associated *only* with cooperative conceptions of resource allocation, in the classical sense that I have discussed.

4.2 Other allocation rules that can be efficiently Kantian implemented

In fact, there are some other allocation rules that can be efficiently implemented as K^+ equilibria. Before discussing this, let's note that I defined the share functions θ^i as being functions of *G* as well as of the effort profile **E**. One might question this generality, because none of the allocation rules X_β do, in fact, have share rules that are a function of *G*. However, it is good not to exclude such rules: for instance, the Walrasian share rule is a function of *G*. To wit, let the firm that operates *G* be owned by agents in shares $\{\sigma^i\}$. The Walrasian share rule is defined by:

$$\theta^{i,Wal}(E^1,...,E^n,G) = \frac{G'(E^S)}{G(E^S)}E^i + \sigma^i(1 - E^S\frac{G'(E^S)}{G(E^S)}) \ .$$

(Just compute that $\theta^{i,Wal}G(E^S)$ equals *i*'s wage income plus her share of profits.) *Not* to allow the shares to depend on *G* would thus eliminate by fiat this important rule.

Fix a concave production function *G*. Let $\theta = (\theta^1,...,\theta^n)$ define a share rule, where the components depend upon **E** and *G*. Then the effort vector **E** is a K^+ equilibrium for the share rule θ at the profile of concave utility functions $u = (u^1,...,u^n)$ only if the following first-order condition holds:

$$(\forall i)(u_1^i \cdot \left((\nabla\theta^i \cdot \mathbf{1})G(E^S) + n\theta^i(\mathbf{E})G'(E^S)\right) + u_2^i) = 0 \,, \qquad (4.13)$$

where **1** is the *n*-vector of one's. This effort allocation is Pareto efficient if and only if the coefficient of u_1^i equals the marginal rate of transformation:

$$\left((\nabla\theta^i \cdot \mathbf{1})G(E^S) + n\theta^i(\mathbf{E})G'(E^S)\right) = G'(E^S),$$

or $$(\forall i) \quad (\nabla\theta^i \cdot \mathbf{1})G(E^S) = (1 - n\theta^i)G'(E^S). \tag{4.14}$$

Now I claim that for (θ, K^+) to be an efficient Kantian pair, equation (4.14) must hold for all vectors $\mathbf{E} \in \Re^n_+$. The reason is that, given any such vector **E**, we can find a profile of utility functions $u = (u^1, ..., u^n)$ such that the marginal rate of substitution of u^i at the point **E** is exactly $\left((\nabla\theta^i \cdot \mathbf{1})G(E^S) + n\theta^i(\mathbf{E})G'(E^S)\right)$, and by equation (4.13), it follows that **E** is a Pareto efficient K^+ equilibrium for the allocation rule θ for the economy (u, G). Therefore equations (4.14) comprise a system of *n* partial differential equations defined on vectors $\mathbf{E} \in \Re^n_+$ that characterize the pair (θ, K^+) being an efficient Kantian pair.

The next step is to reduce this system to one of ordinary differential equations. Let **E** be an arbitrary positive *n*-vector. Consider the system of equations in the single real variable *x*:

$$(\forall i) \quad (\nabla\theta^i(\mathbf{E} + \mathbf{x}) \cdot \mathbf{1})G(E^S + nx) = (1 - n\theta^i(\mathbf{E} + \mathbf{x}))G'(E^S + nx) \tag{4.15}$$

where $\mathbf{x} = (x, x, ..., x) \in \Re^n$. These differential equations must hold for all $x \geq -\min_i E^i$. Now define $\mu^i(x) = \theta^i(E^1 + x, ..., E^n + x)$ and $\rho(x) = G(E^S + nx)$. Note that $(\mu^i)'(x) = \nabla\theta^i(\mathbf{E} + \mathbf{x}) \cdot \mathbf{1}$ and $\rho'(x) = nG'(E^S + nx)$. Therefore the system (4.15) can be written:

$$(\forall i) \quad (\mu^i)'(x)\rho(x) = (1 - n\mu^i(x))\frac{\rho'(x)}{n}, \tag{4.16}$$

which can further be written $$\frac{\rho'(x)}{\rho(x)} = \frac{n(\mu^i)'(x)}{1 - n\mu^i(x)}. \tag{4.17}$$

Now if $1 - n\mu^i(x) > 0$ then the ordinary differential equation (4.17) integrates to:

$$\log\rho(x) + \log(1 - n\mu^i(x)) = \hat{k}, \tag{4.18}$$

where *k* is a constant that may depend upon the vector **E.** In turn, equation (4.18) may be written $\rho(x)(1 - n\mu^i(x)) = k$, where *k* is now a positive constant. But this says that:

$$\theta^i(\mathbf{E}+\mathbf{x}) = \frac{1}{n} - \frac{k(\mathbf{E})}{nG(E^S+nx)} \quad . \tag{4.19}$$

On the other hand, if $1-n\mu^i(x)<0$, then equation (4.17) integrates to:

$$\log\rho(x) + \log(n\mu^i(x)-1) = \tilde{k} \tag{4.20}$$

which in turn can be written $\rho(x)(n\mu^i(x)-1) = k^*$, where k^* is a positive constant. But this says that:

$$\theta^i(\mathbf{E}+\mathbf{x}) = \frac{1}{n} + \frac{k^*(\mathbf{E})}{nG(E^S+nx)} \quad . \tag{4.21}$$

To complete the characterization, it must be the case that $0 \le \theta^i \le 1$ and $\sum_i \theta^i = 1$.

To see what these allocation rules look like, let's suppose that $n=2$. Equations (4.19) and (4.21) say that, if we take any vector $\mathbf{E}=(E^1,E^2)$ then the shares of output along any 45^0 line through $\mathbf{E}$ are given by these two equations and the constants $k^*(\mathbf{E}) = k(\mathbf{E})$. Furthermore it must be the case that $k(\mathbf{E}) \le G(E^S+2x)$ for any $x \ge -\min_i E^i$. Suppose that $E^1 < E^2$. Then the choice of $k = G(E^2-E^1)$ will do. (Here, $x=-E^1$, so $E^S+2x = E^2-E^1$.) We summarize this analysis with:

Proposition 4.2 *The efficient Kantian pairs associated with the additive Kantian variation have allocation rules* $X^i = \theta^i G$ *of the following form:*

$$X^i(\mathbf{E}+\mathbf{x}) = \frac{1}{n}G(E^S+nx) + k^i(\mathbf{E})$$

where $\mathbf{E}$ *may be chosen to be any vector in* $\Re^n_+$ *and* $\mathbf{x}=(x,...,x)$ *where* $x \ge -\min_i E^i$. *The constants* k^i *must have the properties that* $\sum_i k^i(\mathbf{E}) = 0$ *and for all i,*

$$\frac{G(E^S - n\min_j E^j)}{n} + k^i \ge 0 \quad .$$

In this sense, the admissible share rules are those in a neighborhood of the equal-division rule.

A similar analysis shows that the *only* share rule which forms an efficient Kantian pair with the multiplicative Kantian variation is the proportional rule. I will leave the derivation to the reader, providing the clue that to reduce the system of partial differential equations analogous to (4.14) to a system of ordinary ones, one must define the functions $\rho(x) = G(xE^S)$ and $\mu^i(x) = \theta^i(x\mathbf{E})$.

Two final remarks: (1) the Kantian variations φ_β defined in section 4.1 are each associated with a set of allocation rules comprising a neighborhood of the rule X_β , for $0 < \beta < \infty$. It is only the multiplicative variation that is associated with a unique allocation rule. (2) If we restrict the share functions θ^i to depend only on **E** (and not on *G*) then the allocation rules defined in section 4.1 are the unique ones associated with the variations φ_β , in efficient pairs, for $\beta \in [0,\infty]$.

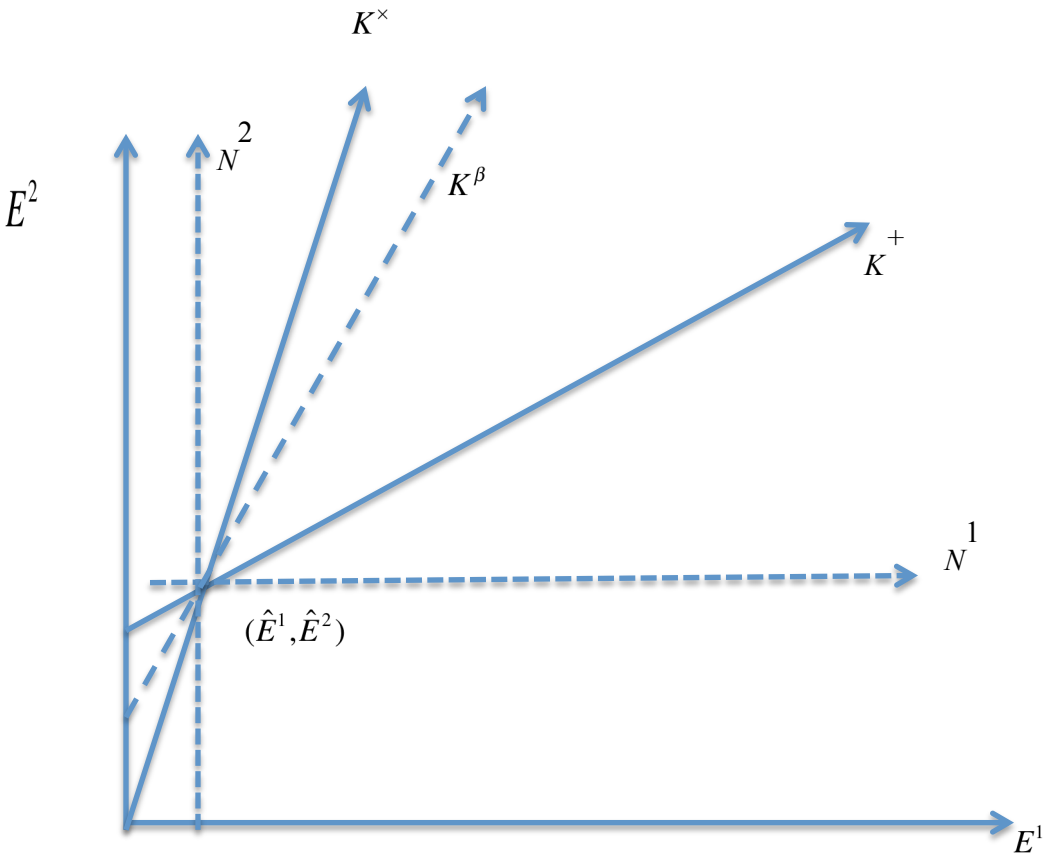

Figure 4.1 Illustration of Kantian counterfactual rays for $n = 2$.

Words: 2161

Chapter 5 Altruism

In this chapter, we study the nature of Kantian equilibria when individuals are altruistic. We will restrict ourselves to examining multiplicative Kantian equilibria in economies $(e, X^{\Pr})$ where e is an environment in the convex domain $\mathfrak{E}$. The results, however, generalize to the kinds of Kantian equilibria we have discussed in chapter 4: that is, K_β equilibria economies (e, X_β) where the Kantian variation is φ_β, $\beta \in [0,\infty]$.

5.1 Altruistic preferences and Pareto efficiency

We will assume that individuals have preferences of the form:

$$U^i(\mathbf{x},\mathbf{E}) = u^i(x^i,E^i) + \alpha^i S(u^1(x^1,E^1),...,u^n(x^n,E^n)) \tag{5.1}$$

where $(\mathbf{x},\mathbf{E})$ is the entire allocation of consumptions and efforts, S is a Bergson-Samuelson social welfare function (which is concave, differentiable and increasing in its arguments), and $\alpha^i \geq 0$. If $\alpha^i = 0$, the individual is self-interested, and if $\alpha^i = \infty$, he is a pure altruist, caring only about social welfare.

We begin by characterizing what the Pareto efficient allocations will be in economies with the 'all-encompassing preferences U^i' and a concave production function G.

Proposition 5.1 *Suppose $\alpha^i = \alpha$ for all i. An interior allocation $(\mathbf{x},\mathbf{E})$ is Pareto efficient in the economy $(\{U^i\},G)$ if and only if:*

$$(\forall i) \quad -\frac{u_2^i(x^i,E^i)}{u_1^i(x^i,E^i)} = G'(E^S) \text{ , and} \tag{5.2}$$

$$\alpha \leq \left(\max_i (u_1^i \cdot S_i \sum_k (u_1^k)^{-1} - \sum_k S_k \right)^{-1} , \tag{5.3}$$

where S_k is the k^{th} partial derivative of S. All functions are evaluated at the allocation $(\mathbf{x},\mathbf{E})$. [1]

The first condition (5.2) simply says that the allocation is Pareto efficient in the economy with self-interested preferences, when $\alpha = 0$. Clearly this is a necessary

[1] See Dufwenberg et al (2011) for a thorough study of efficiency when preferences are other-regarding.

condition for efficiency in the economy with $\alpha > 0$. Suppose the allocation were not efficient in the self-interested economy. Then find a Pareto-dominating allocation. Notice that *S* must increase, because it is an increasing function of the utilities of all members of the society. Therefore U^i will increase for everyone – strictly, if *S* is strictly increasing in its arguments.

Because of altruism, if an allocation gives individuals very unequal utilities, then even if it is efficient in the self-interested economy (i.e., if (5.2) holds), it will fail to be efficient in the $\alpha-$ economy, if α is sufficiently large, because everyone would prefer a redistribution that increases social welfare, even at the cost of a reduction in one's personal utility. This is the consideration that condition (5.3) formalizes.
Define $PE(\alpha;\mathbf{u},G,S)$ as the set of Pareto efficient allocations when the economic environment is $(\mathbf{u},G,S)$ and α is the common altruistic parameter, which we will write as $PE(\alpha)$ when there is no possibility of confusion. Notice that as α increases, condition (5.3) becomes increasingly restrictive. It follows that the sets $PE(\alpha)$ are nested: that is

$$\alpha > \alpha' \Rightarrow PE(\alpha') \subseteq PE(\alpha). \tag{5.4}$$

It therefore follows that $PE(\infty) = \bigcap_{\alpha \geq 0} PE(\alpha)$. Typically, in the purely altruistic economy, where $\alpha = \infty$, there will be a unique allocation maximizing social welfare – indeed, this must be the case if *S* is strictly concave. Therefore, the set of Pareto efficient points shrinks to a single point as α increases when *S* is strictly concave. The intuition is clear. If everyone cares only about social welfare, then any allocation that does not maximize social welfare can be improved upon, from the viewpoint of each individual, by moving to the allocation that does maximize social welfare.

To get more intuition about condition (5.3), consider the case of a quasi-linear economy, where $u^i(x,E) = x - h^i(E)$. Then $u_1^i \equiv 1$. Let $\alpha \to \infty$. Then condition (5.3) becomes:

$$(\forall i) \quad nS_i \leq \sum_k S_k \text{ or } S_i \leq \frac{1}{n}\sum_k S_k \ . \tag{5.5}$$

Summing the last set of inequalities over *i* gives $\sum_i S_i \leq \sum_k S_k$. But this last expression must be an *equality*. It therefore follows that $S_i = \frac{1}{n}\sum_k S_k$ for all *i,* which is to say that for all (i,j), $S_i = S_j$. If *S* is an anonymous social welfare function[2], this implies that

$$(\forall i,j) \quad u^i(x^i,E^i) = u^j(x^j,E^j) \ . \tag{5.6}$$

We have proved:

Proposition 5.2 *If S is an anonymous social welfare function and the individual utility functions are quasi-linear, then the only Pareto efficient point in the purely altruistic economy equalizes all individual utilities.*

Proof of Proposition 5.1:

To characterize Pareto efficiency in the economy $(\mathbf{u},G,n,\alpha)$ we solve the program:

$$\max u^1(x^1,E^1) + \alpha S(u^1(x^1,E^1),...,u^n(x^n,E^n))$$

subject to

$$\sum x^j \leq G(E^S) \qquad (\rho)$$

$$j \geq 2 : u^j(x^j,E^j) + \alpha S(u^1(x^1,E^1),...,u^n(x^n,E^n)) \geq k^j \qquad (\lambda^j)$$

The KT conditions are, letting $\lambda^1 = 1$:

$$\begin{aligned} &(\forall j)(\lambda^j u_1^j + \alpha S_j u_1^j \Lambda = \rho) \\ &(\forall j)(\lambda^j u_2^j + \alpha S_j u_2^j \Lambda = -G'(E^S)\rho) \end{aligned} \tag{5.7}$$

where $\Lambda = \sum \lambda^j$.

Substituting the first of these equations into the second gives:

$$\text{for all } j \quad (G'u_1^j + u_2^j)(\lambda^j + \alpha S_j \Lambda) = 0$$

and so

$$G'u_1^j + u_2^j = 0 \text{ for all } j \tag{5.8}$$

---

[2] An anonymous social welfare function takes the same value for any permutation of its arguments. Informally, it ignores the names of individuals.

since the second term is positive. (Recall that $S_j > 0$.) Equation (5.8) says MRS$^\gamma$ = MRT.

Now (5.7) implies that $\lambda^j + \alpha S_j \Lambda = \frac{\rho}{u_1^j}$. Adding up these equations over j and solving for Λ gives:

$$\Lambda = \rho A \text{ where } A \equiv \frac{\sum_j \frac{1}{u_1^j}}{1 + \alpha \sum_j S_j} .$$

It follows that $u_1^1(1 + \alpha S_1 \Lambda) = \Lambda / A$, and so we solve:

$$\Lambda = \frac{u_1^1}{1/A - \alpha u_1^1 S_1} .$$

By substituting this value into the other equations, we compute that:

$$\frac{\lambda^j}{\Lambda} = \frac{1}{A u_1^j} - \alpha S_j .$$

Consequently our KT non-negativity condition is that:

$$\text{for all } j \quad \frac{1}{A u_1^j S_j} \geq \alpha .$$

Now substitute the expression for A and solve for α , giving:

$$\text{for all } j\text{:} \quad 1 \geq \alpha \left(u_1^j S_j \sum_k \frac{1}{u_1^k} - \sum_k S_k \right) .$$

If at least one of the terms in parentheses is positive, then this condition is equivalent to:

$$\alpha \leq \frac{1}{\max_j \left(u_1^j S_j \sum_k \frac{1}{u_1^k} - \sum_k S_k \right)} . \qquad (5.9)$$

Suppose to the contrary that the parenthetical terms are all non-positive. This means that :

$$\text{for all } j \quad \frac{S_j}{\sum S_k} \leq \frac{1/u_1^j}{\sum 1/u_1^k} .$$

This inequality is of the form $\frac{a_j}{\sum a_k} \leq \frac{b_j}{\sum b_k}$ where all *a*' s and *b*'s are positive. It follows that all the inequalities are *equalities*; for if one is strict, both sums cannot add to one. Therefore in this case we have $\left(u_1^j S_j \sum_k \frac{1}{u_1^k} - \sum_k S_k \right) = 0$ for all *j*, and hence (5.9) is true since the right-hand side is infinite.

Therefore, an interior allocation is Pareto efficient if and only if conditions (5.8) and (5.9) hold. ■

5.2 Kantian equilibrium with altruism

We fix the allocation rule X^{Pr}. The first remark is *there may be no Pareto efficient allocations in* $(\mathbf{u},G,\alpha,X^{\text{Pr}})$ *that can be implemented with the rule* X^{Pr}. Suppose α is very large – say, infinity. Then the unique Pareto efficient allocation in economic environment $(\mathbf{u},G,\infty)$ is the one that maximizes social welfare. But this allocation may not (in general , it *will* not) be a proportional allocation. Consider the quasi-linear example of Proposition 5.2. Assuming *S* is anonymous, the unique maximizer of the social welfare function (and therefore the unique Pareto efficient point in this economy) is the one which maximizes the surplus (this determines the effort vector) *and* distributes output to equalize utilities. This allocation will only, by coincidence, be a proportional allocation. It therefore follows that we cannot expect the Kantian equilibrium of economies $(\mathbf{u},G,\alpha,X^{\text{Pr}})$ to be Pareto efficient (always, with respect to the all-encompassing preferences U^i).

Denote the set of $K^{\times}$ equilibria for the economy $(\mathbf{u},G,\alpha,X^{\text{Pr}})$ by $K^{\times}(\alpha,X^{\text{Pr}})$. We have:

Proposition 5.3[3] For all $\alpha \geq 0$, $K^{\times}(\alpha,X^{\text{Pr}}) = K^{\times}(0,X^{\text{Pr}})$.

Proof:

[3] The same method extends the proposition to any of the efficient Kantian pairs (X_β,φ_β) .

1. An allocation $(\mathbf{x},\mathbf{E})$ is a $K^{\times}$ equilibrium for the economy $(\mathbf{u},G,\alpha,X^{\Pr})$ if and only if:

$$(\forall i)\quad \left.\frac{d}{dr}\right|_{r=1}\left(u^i(\frac{E^i}{E^S}G(rE^S),rE^i)+\alpha S(u^1(\frac{E^1}{E^S}G(rE^S),rE^1),...,u^n(\frac{E^n}{E^S}G(rE^S),rE^n)\right)=0 \quad (5.10)$$

.

Denote $\left.\frac{d}{dr}\right|_{r=1} u^i(\frac{E^i}{E^S}G(rE^S),rE^i)\equiv D_r u^i$. Then (5.10) can be written:

$$(\forall i)\quad D_r u^i+\alpha\sum_k S_k\cdot D_r u^k=0\,, \qquad (5.11)$$

from which it follows that for all i , $D_r u^i=c$, a constant. Substituting this constant into (5.11), we have:

$$c+\alpha c\sum_k S_k=0\ .$$

Since $\sum_k S_j>0$, it immediately follows that $c=0$. But this says that $D_r u^i=0$ for all *i,* which is exactly the condition that the allocation is a $K^{\times}$ equilibrium in the economy $(\mathbf{u},G,0,X^{\Pr})$, proving the claim. ■

Proposition 5.3 says that *the Kantian equilibria for an economy with a positive degree of altruism, with respect to an allocation rule, are identical to the Kantian equilibria for the associated economy with purely self- regarding preferences.* Indeed, the proposition is more general than stated: it is easy to check that different agents can have different values of the altruistic parameter α^i and the proof goes through.

The important consequence of Proposition 5.3 is that Kantian equilibria of economies with and without altruism are observationally equivalent! *If* a community has learned to cooperate in the sense of employing Kantian optimization, we cannot tell by observing the equilibrium whether they hold altruistic preferences or not – at least, with altruism modeled in this way. Their altruism has no impact on what happens in the economy. Although Kantian reasoning can deal quite effectively with many kinds of externality (such as the tragedy of the commons, etc.), it has no bite in addressing

altruism. This result does not depend upon the number of individuals' being large: indeed it holds for economies with two people.

Because, as was noted above, there in general will not exist Pareto efficient allocations that can be implemented by a given rule *X,* in the economy with $\alpha > 0$, we should look at *second-best allocations.*

Definition 5.1 Let $PE^{X^{\mathrm{Pr}}}(\alpha;\mathbf{u},G,S)$ be the allocations that satisfy the rule X^{Pr} in the economy (u,G,S,α) and are not Pareto dominated by any X^{Pr}-implementable allocation. Without confusion, we abbreviate the notation to $PE^{X^{\mathrm{Pr}}}(\alpha)$. This is the set of *second-best allocations* with respect to the allocation rule X^{Pr} .

There are reasons that fishing economies use the proportional allocation rule -- because it implements the simple principle 'each fisher keeps his catch.' Likewise, the equal-division allocation rule may be a good rule in hunting societies. So one should ask: What are the best allocations that can be found, *given* that a society is using a particular rule *X*? The second-best allocations $PE^{X}(\alpha)$ comprise good candidates, from the efficiency viewpoint.

We now state:

Proposition 5.4 $PE^{X^{\mathrm{Pr}}}(\alpha) \subset K^{\times}(\alpha,X^{\mathrm{Pr}}) = K^{\times}(0,X^{\mathrm{Pr}})$

In words, the second-best allocations in an altruistic economy, with respect to the proportional allocation rule, are all multiplicative Kantian equilibria in that economy. The last equation in the statement is simply a restatement of Proposition 5.3.

The converse, however, is not true: that is there may be Kantian equilibria which are not in $PE^{X^{\mathrm{Pr}}}(\alpha;\mathbf{u},G)$. Why does the converse not hold? Suppose there are several allocations in $K^{\times}(0,X^{\mathrm{Pr}}) = K^{\times}(\alpha,X^{\mathrm{Pr}})$. Generically, they will be strictly ranked by the social welfare function *S*. So if $\alpha = \infty$, only one of them will be a member of $PE^{X^{\mathrm{Pr}}}(\infty)$. The same argument applies for large finite α because the set $PE^{X^{\mathrm{Pr}}}(\alpha)$ shrinks to a

singleton as α becomes large. Of course, if $K^{\times}(0, X^{\Pr})$ is a singleton, then the converse to Proposition 5.4 does hold.

While Proposition 5.3 is, in a sense, disappointing, because it says that Kantian optimization is incapable of taking into account the altruism that people may feel, Proposition 5.4 is optimistic, because it says that *if* an allocation is second-best Pareto efficient in an economy with altruism, then it *can* be supported as a Kantian equilibrium. Because there is often good reason to view the allocation rule as fixed (as in hunting and fishing economies), second-best Pareto efficiency becomes the appropriate welfare criterion.

Proof of Proposition 5.4:

1. We first observe that if an allocation is in $PE^{X^{\Pr}}(\alpha)$, it must be Pareto efficient in the (self-interested) 0-economy. To show this, we need only show that $MRS^i = MRT$ for all agents i, where the marginal rate of substitution is computed with the personal utilities functions u^i. Suppose, to the contrary, that for some i, $MRS^i < G'(E^S)$ at the allocation $X^{\Pr}(E^1,...,E^n)$. Now consider a small positive increase ε in E^i, holding all other efforts fixed, and look at the new allocation determined by the proportional rule $X^{\Pr}$. The *personal* utility of i will increase if :

$$\frac{d}{dE^i} u^i\left(\frac{E^i}{E^S} G(E^S), E^i\right) > 0 \ . \tag{5.12}$$

This derivative evaluates to:

$$u_1^i\left(\frac{E^i}{E^S} G'(E^S) + \frac{E^S - E^i}{(E^S)^2} G(E^S)\right) + u_2^i >^? 0 \ .$$

Now use the fact that $u_1^i G' + u_2^i > 0$ (that is, $MRS^i < G'$) and the desired inequality will follow if:

$$-u_2^i \frac{E^i}{E^S} + u_1^i \frac{E^S - E^i}{(E^S)^2} G(E^S) >^? -u_2^i$$

which can be rewritten as:

$$-\left(\frac{u_2^i}{u_1^i}\right)\left(\frac{E^i - E^S}{E^S}\right) + \frac{E^S - E^i}{(E^S)^2} G(E^S) >^? 0 \ .$$

Dividing by the positive number $(E^S - E^i)/E^S$, this inequality reduces to:

$$\frac{G(E^S)}{E^S} >^{?} MRS^i,$$

which is true, because the concavity of *G* implies that $\frac{G(E^S)}{E^S} > G'(E^S)$. This proves (5.12).

2. Now this small increase in i's effort also *increases* the utilities of all other agents to the first-order, because they care about person *i* via S; it also *decreases* their utility, but only to the second order, because there is a second order decrease in G' and hence in the consumptions of the others. Net, the all-encompassing utilities of all the other agents increase. However, because the consumption of the other agents decrease, *i*'s all-encompassing utility *decreases*, because she cares about the others, but this decrease is only to the second order. Therefore, to the first order, all all-encompassing utilities increase, and this contradicts the assumption that the original allocation was in $PE^{X^{\Pr}}(\alpha)$.

3. It follows that it must be that $MRS^i = MRT$ for all agents i and hence the allocation is 0-Pareto efficient. We can state this fact as : $PE^{X^{\Pr}}(\alpha) \subseteq PE^{X^{\Pr}}(0)$.

4. Denote by $X^{\Pr}[G]$ the set of all proportional allocations that are feasible given the production function *G*. We know that $K^{\times}(0, X^{\Pr}) = X^{\Pr}[G] \cap PE^{X^{\Pr}}(0)$. That is, multiplicative Kantian equilibria for the proportional allocation rule are precisely the positive, proportional allocations that are Pareto efficient (in the self-interested economy). By virtue of Proposition 5.3, we have $K^{\times}(\alpha, X^{\Pr}) = X^{\Pr}[\mathbf{u}, G] \cap PE(0)$. By virtue of step 3 of this proof, we therefore have

$PE^{X^{\Pr}}(\alpha) = X^{\Pr}[G] \cap PE^{X^{\Pr}}(\alpha) \subset X^{\Pr}[G] \cap PE^{X^{\Pr}}(0) = K^{\times}(\alpha, X^{\Pr})$, proving the proposition. ■

5.3 Review

If members of a community are altruistic towards each other, and if each is applying the Kantian protocol, the observed outcome is observationally equivalent to what it would be if each had self-regarding preferences. In this sense, altruism is a

gratuitous assumption, which does not appear to have much impact if cooperation is already present. Of course, we cannot expect this result to generalize to other ways of modeling altruism: here, we have taken the traditional approach of appending a social welfare function to each individual's self-interested utility function.

Because of this result, it follows that Kantian equilibria in the presence of altruism will not in general be Pareto efficient (with respect to the altruistic preferences) – Kantian optimization will not take account of the consumption externalities in agents' preferences.

Nevertheless, all second-best Pareto efficient allocations will be Kantian equilibria in the presence of altruism. This is as much as we can ask for.

Words: 3956

Chapter 6 Can one rationalize Kantian optimization as Nash optimization where players have extended preferences?

An important research program of behavioral economics is to attempt to explain phenomena that appear to be equilibria in game situations, but are not Nash equilibria of the game with traditional self-interested preferences[1]. I have characterized the program of behavioral economics as including non-traditional arguments in the preferences of players, so that the Nash equilibrium of the game played by players with these modified preferences is indeed the observed outcome. In this chapter, I pursue the *reverse* question as it applies to Kantian equilibrium. Is it the case that the Kantian equilibrium of a game where players have traditional self-interested preferences is indeed the *Nash* equilibrium of game whose players have modified preferences? If this were the case, then someone skeptical of my approach could say, "Well, you see, what we are observing is just a Nash equilibrium, but where players have modified preferences. Kantian optimization is just a gratuitous diversion. " The implication is there is no need for invoking a new optimization protocol, as I have been urging, to explain cooperation.

For most of this chapter, I will work with the economic environments with production $\mathfrak{E}$ studied in chapter 3. Indeed, I will assume that the allocation rule $X^{\Pr}$ is in place, and the Kantian protocol that players use is $K^{\times}$. Thus, any positive Kantian equilibrium of such an economy is Pareto efficient.

I begin by examining two cases in which the answer to the question posed is easily seen to be affirmative. The first case is when players have quasi-linear preferences; thus $u^i(x^i,E^i) = x^i - h^i(E^i)$ where $\{h^i\}$ are convex functions. Let us suppose $n=2$ to keep things notationally simple. We modify the players' utility functions as follows. Both players will have the *same* extended utility function V given by:

$$V(u^1,u^2) = u^1 + u^2 \ . \tag{6.1}$$

---

[1] Of course, the original path-breaking contribution in behavioral economics, of Kahneman and Tversky (1978), was concerned not with games, but with optimization in decision problems.

It is important to realize that in (6.1), each player has preferences over the whole allocation (x^1,E^1,x^2,E^2) . Thus, we should write out (6.1) in full:

$$V(u^1(x^1,E^1),u^2(x^2,E^2)) = u^1(x^1,E^1)+u^2(x^2,E^2)\,. \tag{6.2}$$

Formally, $V:\Re^4_+ \to \Re$.

Let us write the payoff functions of the game where both players have the utility function *V*, where the strategies are effort levels:

$$\hat{V}(E^1,E^2) = \frac{E^1}{E^S}G(E^S) - h^1(E^1) + \frac{E^2}{E^S}G(E^S) - h^2(E^2)\ ; \tag{6.3}$$

here, we have just used the proportional allocation rule and substituted into (6.1). Although the players have the same preferences, the game is not symmetric with respect to the strategies of the two players. As always, player one controls E^1 and player two controls E^2. Now we look at the *Nash* equilibrium of this new game. The first-order conditions for Nash equilibrium are:

$$\frac{\partial \hat{V}}{\partial E^1} = 0 \text{ and } \frac{\partial \hat{V}}{\partial E^2} = 0\ . \tag{6.4}$$

Compute that these conditions are exactly:

$$G'(E^S) = (h^1)'(E^1) \text{ and } G'(E^S) = (h^2)'(E^2)\ . \tag{6.5}$$

But these are precisely the conditions for Pareto efficiency! Therefore the Nash equilibrium of the game is a proportional allocation that is Pareto efficient: this is precisely the $K^\times$ equilibrium of the game with players whose preferences are given by (u^1,u^2) .

In other words, the Kantian equilibrium of the game with self-interested (quasi-linear) preferences is exactly the Nash equilibrium of the game with players each of whom is concerned with maximizing the *sum* of utilities of the players. The two games are, in a word, observationally equivalent. There is, however, one drawback to this example: it requires a cardinal representation of the utility functions. For instance, if we used the utility function $2u^1$ for the first player, and continue to use u^2 for the second player, and define the extended preferences as $V = 2u^1 + u^2$, it is no longer the case that the Nash equilibrium of the new game is Pareto efficient. Indeed, with this *V*, the first-order condition for Nash equilibrium for the first player is:

$$\frac{G(E^S)}{E^S}\left(\frac{E^2}{E^S}\right)+G'(E^S)\left(1+\frac{E^1}{E^S}\right)=(h^1)'(E^1)\ . \tag{6.6}$$

The second case I propose is when all players have the *same* utility function *u*. Again, we define the extended preferences as the 'altruistic' ones:

$$V(u,u)=u[1]+u[2]\ , \tag{6.7}$$

where the notation $u[i]$ means the utility of player *i* at the allocation.

The players now have the symmetric payoff function:

$$\hat{V}(E^1,E^2)=u(\frac{E^1}{E^S}G(E^S),E^1)+u(\frac{E^2}{E^S}G(E^S),E^2)\ . \tag{6.8}$$

The condition for Nash equilibrium is again (6.4). These conditions compute to:

$$\begin{aligned} &u_1[1]\left(\frac{E^1}{E^S}(G'(E^S)-\frac{G(E^S)}{E^S})+\frac{G(E^S)}{E^S}\right)+u_2[1]+u_1[2]\left(\frac{E^2}{E^S}(G'(E^S)-\frac{G(E^S)}{E^S})\right)=0\\ &u_1[2]\left(\frac{E^2}{E^S}(G'(E^S)-\frac{G(E^S)}{E^S})+\frac{G(E^S)}{E^S}\right)+u_2[2]+u_1[1]\left(\frac{E^1}{E^S}(G'(E^S)-\frac{G(E^S)}{E^S})\right)=0 \end{aligned}\ . \tag{6.9}$$

Let us look for a symmetric solution to equations (6.9). At such a allocation, we have $u_1[1]=u_1[2]$ and $u_2[1]=u_2[2]$. Then the first equation becomes:

$$u_1[1]\left(\frac{E^S}{E^S}(G'(E^S)-\frac{G(E^S)}{E^S}\right)+u_1[1]\frac{G(E^S)}{E^S}+u_2[1]=0 \text{ or } u_1[1]G'(E^S)+u_2[1]=0\ , \tag{6.10}$$

and the second equation similarly reduces to $u_1[2]G'(E^S)+u_2[2]=0$. But these two equations simply say that the marginal rate of transformation is equal to the marginal rate of substitution for both players. In other words, the multiplicative Kantian equilibrium (which is a symmetric, proportional allocation that is Pareto efficient) solves (6.7), and is therefore a Nash equilibrium of the game played by players with extended preferences.

Again, the same drawback applies: we must use cardinal utility functions that are identical for the two players, for this conclusion to hold. We have shown:

Proposition 6.1

A. *Suppose players have quasi-linear utility functions of the form* $u^i(x^i, L^i) = x^i - h^i(E^i)$, or *suppose all players have the same concave utility function u. Consider the new game where each player maximizes the sum of all the (cardinal) utility functions of players in the original game. Then there is a Nash equilibrium of the game with extended preferences which is the* $K^{\times}$ *of the original game (with self-interested preferences).*

B. *If we substitute* K^{+} *for* $K^{\times}$ *in statement* A*, the statement is still true.*

Proof of part B replicates the proof given for part A.

It turns out that the result with identical utility functions extends to many other games. In the next proposition, we consider games in which player *i*'s payoff function is $P(E^i, E^S)$ for some function *P*. This covers symmetric public-good games (so-called trust games in lab experiments) and all two-person symmetric games.

Proposition 6.2 *Define the extended preferences* $V(E^1, E^2, ..., E^n) = \sum_i P(E^i, E^S)$. *If* $\mathbf{E} = (E, ..., E)$ *is an interior, symmetric* $K^{\times}$ *equilibrium of the game where players' payoff functions are* $P(E^i, E^S)$, *then it is a Nash equilibrium of the game where all players have the 'extended' payoff function V.*

Proof:

If $\mathbf{E}$ is a symmetric interior $K^{\times}$ equilibrium of the first game then the first-order conditions

$$(\forall i) \quad \left.\frac{d}{dr}\right|_{r=1} P(rE, rE^S) = 0 \tag{6.11}$$

or

$$(\forall i) \quad P_1[i]E + P_2[i]nE = 0 \tag{6.12}$$

where $[i]$ means the function is evaluated at (E^i, E^S). But the symmetry assumption means that (6.12) reduces to

$$(\forall i) \quad P_1[i] + nP_2[i] = 0 \ . \tag{6.13}$$

The f.o.c.s for Nash equilibrium of the game with extended preferences are:

$$(\forall i) \quad P_1[i] + \sum_{j=1}^{n} P_2[j] = 0. \tag{6.14}$$

Since the Kantian allocation is symmetric, we have for all j, $P_2[j] = P_2[1]$, and hence (6.14) reduces to (6.13), proving the claim. ■

Many laboratory experiments have set-ups in which participants are equipped with symmetric preferences of the form $P(E^i, E^S)$. Proposition 6.2 asserts that we cannot tell, if people achieve the cooperative solution, whether they are using the Kantian protocol with their self-interested preferences, or the Nash protocol with a utilitarian form of altruism – each maximizing the total payoff.

We now ask whether these examples – of quasi-linear utility functions, or identical utility functions – can be extended. Fix a pair of concave utility functions (u^1, u^2) for the production economies. Consider the domain of economies $\Omega^{(u^1,u^2)} = \{(u^1, u^2, G, X^{\text{Pr}}) \mid G \text{ is non-linear, concave}\}$. Any $K^\times$ equilibrium for an economy in this domain generates a pair of utilities for the players, say (a,b). Let $D^{(u^1,u^2)} \subset \Re^2$ be the set of such utility pairs. The next two propositions show that, indeed, we can 'rationalize' $K^\times$ equilibria on $\Omega^{(u^1,u^2)}$ as Nash equilibria on economies where both players have extended utility functions over the domain of allocations in $\Re^4_+$.

Proposition 6.3 *Suppose there are extended utility functions* $V^1(u^1, u^2)$ and $V^2(u^1, u^2)$ *such that a Nash equilibrium of the game* $(V^1, V^2, G, X^{\text{Pr}})$ *is a* $K^\times$ *equilibrium of the game* $e = (u^1, u^2, G, X^{\text{Pr}})$ *for any* $e \in \mathfrak{E}$ *where G is non-linear. Then* V^1 and V^2 *represent the same ordinal preferences [i.e., have the same indifference map] on the set* $D^{(u^1,u^2)}$.

Let's explain what this says. V^1 and V^2 are defined on allocations in $\Re^4_+$. However, they can also be viewed as having indifference curves in $\Re^2_+$, since any allocation produces a pair of utility numbers. The proposition claims that if a Kantian equilibrium can be rationalized as a Nash equilibrium with extended preferences on the domain $\Omega^{(u^1,u^2)}$, those preferences must be the same for both players. So the

constructions of Proposition 6.1 -- where both players had the same extended preference order over utility pairs – was not a coincidence.

Proof:

1. We are given a pair of extended utility functions $V^1(u^1,u^2)$ and $V^2(u^1,u^2)$. As before, we define the associated payoff functions of game with the proportional allocation rule by:

$$\text{for } i=1,2: \quad \hat{V}^i(E^1,E^2)=V^i(u^1(\frac{E^1}{E^S}G(E^S),E^1),u^2(\frac{E^2}{E^S}G(E^S),E^2)) \ . \tag{6.15}$$

The f.o.c.s for a Nash equilibrium of the extended games are:

$$\frac{\partial\hat{V}^1(E^1,E^2)}{\partial E^1}=\frac{\partial\hat{V}^2(E^1,E^2)}{\partial E^2}=0 \ . \tag{6.16}$$

Writing out the derivative[2] for $\hat{V}^1$, using (6.15), we have:

$$\left(V_1^1u_1^1\frac{E^1}{E^S}+V_2^1u_1^2\frac{E^2}{E^S}\right)G'(E^S)+\frac{G(E^S)}{E^S}\frac{E^2}{E^S}(V_1^1u_1^1-V_2^1u_1^2)+V_1^1u_2^1=0 \, . \tag{6.17}$$

Dividing by u_1^1 (which is positive) and using the fact that $-G'(E^S)=\frac{u_2^1}{u_1^1}$, since by hypothesis the Nash equilibrium is the multiplicative Kantian equilibrium, and is hence Pareto efficient, we have:

$$\begin{aligned}
&\left(V_1^1\frac{E^1}{E^S}+V_2^1\frac{u_1^2}{u_1^1}\frac{E^2}{E^S}\right)G'(E^S)+\frac{G(E^S)}{E^S}\frac{E^2}{E^S}(V_1^1-V_2^1\frac{u_1^2}{u_1^1})-V_1^1G'(E^S)=0\\
&G'(E^S)\left(V_1^1(\frac{E^1}{E^S}-1)+V_2^1\frac{u_1^2}{u_1^1}\frac{E^2}{E^S}\right)+\frac{G(E^S)E^2}{(E^S)^2}\left(V_1^1-\hat{V}_2^1\frac{u_1^2}{u_1^1}\right)=0\\
&G'(E^S)\frac{E^2}{E^S}(-V_1^1+V_2^1\frac{u_1^2}{u_1^1})+\frac{G(E^S)E^2}{(E^S)^2}\left(V_1^1-V_2^1\frac{u_1^2}{u_1^1}\right)=0
\end{aligned} \tag{6.18}$$

from which it follows that either $V_1^1=V_2^1\frac{u_1^2}{u_1^1}$ or $G'=\frac{G(E^S)}{E^S}$. But the second possibility is false because *G* is concave but non-linear. Therefore we must have:

---

[2] V_j^1 is the derivative of V^1 with respect to the utility of the *j*th player.

$$V_1^1 = V_2^1 \frac{u_1^2}{u_1^1} \tag{6.19}$$

at the proportional solution on the whole domain of economic environments. In like manner, we can expand the second condition in (6.16) to give:

$$V_2^2 = V_1^2 \frac{u_1^1}{u_1^2} \tag{6.20}$$

on the whole domain. Therefore, $\frac{V_2^2}{V_1^2} = \frac{V_2^1}{V_1^1}$ at all proportional solutions on the domain $D^{(u^1,u^2)}$.

2. Therefore $\frac{V_2^2}{V_1^2} = \frac{V_2^1}{V_1^1}$ is an identity on $D^{(u^1,u^2)}$. But this means that the *marginal rates of substitution* of V^1 and V^2, viewed now as functions on $\Re^2$, are identical on domain $D^{(u^1,u^2)}$, and hence V^1 and V^2 have identical indifference maps in $\Re_+^2$. Hence they represent the *same preferences* over pairs of utilities, proving the claim. ■

Because Kantian and Nash equilibrium are both concepts defined on *preferences*, this means that we can assume that if Kantian equilibrium can be everywhere rationalized as a Nash equilibrium with extended preferences, those extended preferences (over pairs of self-interested utilities) are the same for the two players. In other words, if such a representation exists, we may take $V^1 = V = V^2$, for some function $V(u^1,u^2)$.

I will now seek to construct the extended utility function V for a general Cobb-Douglas economy. Let $u^1(x,E) = x(1-E)^m, \quad u^2(x,E) = x(1-E)^n, \quad 0 < m < n < \infty$. These are two different Cobb-Douglas utility functions. On the domain $\Omega^{(u^1,u^2)}$ we know that multiplicative Kantian equilibrium is Pareto efficient. We know that Nash equilibrium is defined and is always Pareto inefficient. We study whether it is possible to construct extended preferences V for each player such that the Nash equilibrium on the extended economy is the Kantian equilibrium for $(u^1,u^2,G,X^{\Pr})$.

Note that if *G* were linear, then the Nash equilibria on $\Omega^{(u^1,u^2)}$ are efficient and they therefore coincide with the multiplicative Kantian equilibria. I have eliminated this case from the domain.

Eqn. (6.19) tells us how the slopes of the indifference curves of V^i are related to the Kantian equilibrium. Now (6.19) says that it must be the case that:

$$\frac{V_2(a,b)}{V_1(a,b)} = \frac{(1-E^1)^m}{(1-E^2)^n} \tag{6.21}$$

where $a = u^1(x^1,E^1), b = u^2(x^2,E^2)$ at the Nash allocation (which is also the Kantian allocation of the original economy). To prove the affirmative claim, we must show that there exists a function $\Phi : \Re^2_+ \to \Re$ such that at any Kantian equilibrium on the domain $\Omega^{(u^1,u^2)}$, it is true that $\Phi(a,b) = -\frac{(1-E^1)^m}{(1-E^2)^n}$ where (E^1,E^2) is the Kantian effort vector which yields utilities *a,b*. For if this is true, then we have characterized $-\frac{V_2}{V_1}$ on its domain (namely: $-\frac{V_2^1(a,b)}{V_1^1(a,b)} = \Phi(a,b)$) and so (we will later argue) have characterized the indifference map of V. The proof will establish the existence of such a function Φ and then of *V,* which will as well be differentiable.

We next observe that an interior Kantian allocation on the domain $\Omega^{(u^1,u^2)}$ is characterized by the following two equations and inequality:

$$\frac{mx^1}{1-E^1} = \frac{nx^2}{1-E^2} \quad , \tag{6.22}$$

$$\frac{x^1}{E^1} = \frac{x^2}{E^2} \ , \tag{6.23}$$

and

$$m\frac{E^1}{1-E^1} < 1 \tag{6.24}$$

Eqn. (6.22) says the MRSs of the two players are equal. (6.23) says the allocation is proportional. (6.24) is equivalent to $m\frac{x^1}{1-E^1} < \frac{x^1}{E^1} = \frac{x^1+x^2}{E^1+E^2}$. We can therefore find a strictly concave *G* whose slope at the point (E^1+E^2, x^1+x^2) equals the MRS, because the last inequality tells us the marginal product of this *G* is less than its average product at this point, which is the condition for finding a nonlinear concave *G* passing through the point. Indeed, note that (6.24) can be rewritten as:

$$E^1 < \frac{1}{m+1} \quad . \tag{6.25}$$

We now re-write the equations characterizing the interior Kantian equilibria on the domain $\Omega^{(u^1,u^2)}$ as follows:

$$\begin{aligned} &mE^1 - nE^2 + (n-m)E^1E^2 = 0 \\ &x^1E^2 - x^2E^1 = 0 \\ &x^1(1-E^1)^m = a \\ &x^2(1-E^2)^n = b \\ &E^1 < \frac{1}{m+1} \end{aligned} \tag{6.26}$$

where we have included the utility values as well. We can view these equations as ones defining the entire class of interior Kantian allocations on $\Omega^{(u^1,u^2)}$, where E^1 is restricted to the interval $(0, \frac{1}{m+1})$, as a function of the two parameters (a,b) .

Our procedure will be to show, using the implicit function theorem, that the four variables E^1, E^2, x^1, x^2 can be defined as (differentiable) functions of (a,b) on the solution space of (6.26). It will then immediately follow that we have constructed the function $\Phi(a,b) = -\frac{(1-E^1(a,b))^m}{(1-E^2(a,b))^n}$, which will complete the proof. (Note the denominator in the definition of Φ is never zero, because $E^2 < \frac{1}{n+1}$.)

To show this, we will demonstrate that the Jacobian of the system (6.26) never vanishes on the domain $D^{(u^1,u^2)}$. Order the variables (x^1,E^1,x^2,E^2) and compute that the Jacobian of (6.26) is:

$$J=\begin{pmatrix} 0 & m+(n-m)E^2 & 0 & -n+(n-m)E^1 \\ E^2 & -x^2 & -E^1 & x^1 \\ (1-E^1)^m & -mx^1(1-E^1)^{m-1} & 0 & 0 \\ 0 & 0 & (1-E^2)^n & -nx^2(1-E^2)^{n-1} \end{pmatrix}. \qquad (6.27)$$

Expanding the determinant of J and dividing by the positive number $(1-E^1)^{m-1}(1-E^2)^{n-1}$ demonstrates that $|J|\neq 0$ if and only if:

$$\begin{aligned} &\overbrace{((n-m)E^1-n)}^{neg}(1-E^2)x^2(1-E^1)\overbrace{(\frac{mE^1}{1-E^1}-1)}^{neg}+ \\ &(1-E^1)\overbrace{(m+(n-m)E^2)}^{pos}x^1(1-E^2)\overbrace{(\frac{nE^2}{1-E^2}-1)}^{neg}\neq 0 \end{aligned} \qquad (6.28)$$

Suppose $|J|=0$. By (6.22), $(\frac{mE^1}{1-E^1}-1)=(\frac{nE^2}{1-E^2}-1)$, and so we can rewrite the *negation* of (6.28) as:

$$\overbrace{((n-m)E^1-n)}^{neg}(1-E^2)E^2(1-E^1)+(1-E^1)\overbrace{(m+(n-m)E^2)}^{pos}E^1(1-E^2)=0$$

and further simplify to:

$$\overbrace{((n-m)E^1-n)}^{neg}E^2+\overbrace{(m+(n-m)E^2)}^{pos}E^1=0$$

which in turn simplifies to:

$$nE^2=mE^1 \ . \qquad (6.29)$$

But this contradicts the fact that $\frac{mE^1}{1-E^1} = \frac{nE^2}{1-E^2}$, since $m \neq n$, which proves that $|J| \neq 0$.

It therefore follows by the implicit function theorem that there is a differentiable function Φ that can be defined locally around any point (a,b). What is the relevant global extension of this result? We have to consider the domain $D^{(u^1,u^2)} \in \Re^2$ for which there exists a non-linear concave differentiable production function G at which the multiplicative Kantian equilibrium of the economy $(u^1,u^2,G,X^{\Pr})$ is a point $(a,b) \in D^{(u^1,u^2)}$. Our existence results (proved in chapter 7 below) will show this is a large domain. The global inverse theorem of Hadamard[3] allows us to extend the locally defined function Φ to any compact subset of D. (We will not check this here.) Except for checking the Hadamard extension, we have therefore proved that there exists a continuously differentiable function $\Phi(a,b)$ which specifies the marginal rate of substitution of the social welfare function $V(a,b)$, whose existence we wish to prove, at every point (a,b) for which the system (6.26) can be solved.

We must finally ask: Can we indeed 'integrate' the function Φ to find the function V? The answer is yes. Consider the differential equation:

$$\frac{da}{db} = \Phi(a,b) \ . \tag{6.30}$$

Because Φ is continuously differentiable, by the Picard-Lindelöf theorem, there is a unique solution to the differential equation (6.30) for any initial condition on the function

---

[3] Theorem (Hadamard) Let $F : M_1 \to M_2$ be a continuous function between two smooth, connected manifolds of $\Re^n$. Suppose that:

1. F is proper
2. The Jacobian is everywhere invertible
3. M_2 is simply connected

Then F is a homeomorphism (and hence globally bijective).

a. Denote the solution by $Q(a,b,k)=0$, where *k* is a constant associated with the initial condition. Now differentiating this equation with respect to *b* gives:

$$Q_1\frac{da}{db}+Q_2=0 \text{ or } \frac{da}{db}=\Phi(a,b)=-\frac{Q_2}{Q_1}\ .$$

But this means that the $Q(a,b,k)=0$ is the locus of the *k*th indifference curve of the function *V*. By varying *k*, we sweep out the entire indifference map of *V*. The uniqueness guaranteed by the Picard-Lindelöf theorem tells us that the function *V* is unique, up to ordinal transformation, which will preserve the indifference map[4].

Proposition 6.4 *Let* $u^1(x,E)=x(1-E)^m,\quad u^2(x,E)=x(1-E)^n,\quad 0<m<n<\infty$. *Then there exists a differentiable function* $V:\Re^4\to\Re$ *, such that in a game induced by the economy* $(V,V,G,X^{\Pr})$ *where G is any non-linear concave differentiable production function, and where the preferences of each player are given by* $V(u^1(\cdot,\cdot),u^2(\cdot,\cdot))$ *, the Nash equilibrium is the multiplicative Kantian equilibrium of the game induced by* $(u^1,u^2,G,X^{\Pr})$ *. Furthermore, if* $(V^1,V^2,G,X^{\Pr})$ *is any game with extended preferences whose Nash equilibrium is the multiplicative Kantian equilibrium of* $(u^1,u^2,G,X^{\Pr})$, *then* V^i *is ordinally equivalent to V for* $i=1,2$ *.*

It is interesting to note where strict concavity of *G* enters. If a Kantian equilibrium is associated with a linear *G*, then (6.24) becomes an *equality*. This in turn means that the expression in (6.28) becomes *zero,* because $(\frac{mE^1}{1-E^1}-1)=(\frac{nE^2}{1-E^2}-1)=0$. Therefore $|J|=0$. Note also in Proposition 6.3, I used that fact that $G'(E^S)\neq\frac{G(E^S)}{E^S}$ to deduce eqn. (6.19), with which the proof of that proposition begins.

Let us note why it is easy to construct the extended preferences in the quasi-linear case and in the case when the players have the same utility function. In these two cases,

---

[4] I am grateful to Burak Ünveren for the last paragraph in this proof.

the fundamental equation (6.19) becomes $-\frac{V_2}{V_1}=-1$. This means $V(a,b)=a+b$,

which gives another proof of proposition 6.1.

It is indeed possible to extend Proposition 6.3 to any pair of fixed concave, differentiable utility functions (u^1,u^2) . We begin by replacing the system (6.22)-(6.24) with the general system:

$$\begin{aligned}
&\frac{u_2^1(x^1,E^1)}{u_1^1(x^1,E^1)}=\frac{u_2^2(x^2,E^2)}{u_1^2(x^2,E^2)}\\
&x^1E^2-x^2E^1=0\\
&u^1(x^1,E^1)=a\\
&u^2(x^2,E^2)=b\\
&-\frac{u_2^1(x^1,E^1)}{u_1^1(x^2,E^2)}<\frac{x^1}{E^1}
\end{aligned}\tag{6.31}$$

and then compute its Jacobian, and so on. In general, the extended function V will not be computable in closed form.

The next question I ask is what happens to the function V if we replace $\{u^i\}$ with ordinal transformations of them. Let $\hat{u}^1=f\circ u^1$ and $\hat{u}^2=g\circ u^1$ where f and g are strictly monotone increasing functions. Consider the analogous system to (6.31):

$$\begin{aligned}
&\frac{f'(a)u_2^1(x^1,E^1)}{f'(a)u_1^1(x^1,E^1)}=\frac{g'(b)u_2^2(x^2,E^2)}{g'(b)u_1^2(x^2,E^2)}\\
&x^1E^2-x^2E^1=0\\
&f(u^1(x^1,E^1))=f(a)\\
&g(u^2(x^2,E^2))=f(b)\\
&-\frac{f'(a)u_2^1(x^1,E^1)}{f'(a)u_1^1(x^2,E^2)}<\frac{x^1}{E^1}
\end{aligned}\tag{6.32}$$

A quick comparison shows this is *identical* to the original system (6.31) – just apply f^{-1} and g^{-1} to the third and fourth equations. Let the extended utility function for the new system be denoted by $\hat{V}$ and the extended preferences for the system (6.31) be denoted by V. Then we have:

$$\hat{V}(f(a),g(b)) = V(a,b), \tag{6.33}$$

or, writing this slightly differently:

$$\hat{V}(f(u^1),g(u^2)) = V(u^1,u^2) \tag{6.34}$$

This means that $\hat{V}$ and V define the *same preferences* on arguments $(x^1,E^1,x^2,E^2) \in \Re^4_+$.

Let us now move from the language of utility functions to the language of preference orders. Denote the set of self-interested preference orders over consumption and labor by **R**. Denote the set of preference orders over the whole allocation for two players (an element in $\Re^4_+$) by **Q**. The correct way of stating the central question of this chapter is: given concavifiable preference orders $R^1,R^2 \in \mathbf{R}$, is there a mapping $F:\mathbf{R}^2 \to \mathbf{Q}$ such that the $K^\times$ equilibrium of the economy $(R^1,R^2,G,X^{\text{Pr}})$ is always a Nash equilibrium of the economy $(F(R^1,R^2),F(R^2,R^1),G,X^{\text{Pr}})$? F, in this case, is a social choice rule: given any two preference orders, it aggregates them into a preference order for society. Because the utility function is a derived concept, we must state the query using the fundamental notion of preference orders. We have:

Proposition 6.5 *There exists a social choice rule* $F:\mathbf{R}^2 \to \mathbf{Q}$ *such that for any pair of self-interested preferences* $R^1,R^2 \in \mathbf{R}$*, the* $K^\times$ *equilibria of the economy* $(R^1,R^2,G,X^{\text{Pr}})$ *is a Nash equilibrium of the economy with extended preferences* $(F(R^1,R^2),F(R^2,R^1),G,X^{\text{Pr}})$ *, for all increasing, concave production functions G.*

The proof is, in fact, the equation (6.34). For this equation says that the induced preferences on $\Re^4_+$ (for which the Nash equilibrium is the Kantian equilibrium of the problem with preference orders on $\Re^2_+$) are independent of the utility functions that are chosen to represent the self-regarding preferences. In other words, the induced preferences on the whole allocation that both players have in the 'extended' game depend only on their *preferences* in the original game.

At this point, our skeptic can say: "Well, you see, Kantian optimization really isn't any different from Nash optimization. One just has to realize that people have

preferences over the whole allocation." Proposition 6.5 could be taken to vindicate the program of behavioral economics.

But I strongly demur. The reason is that the function V that must be constructed, that represents the preferences of the players on the whole allocation, is in general extremely complex. Let's take a *simple* example to illustrate. Suppose that the two players in a production game have utility functions:

$$\hat{u}^1(x^1,E^1) = (x^1 - h^1(E^1))^{1/3}, \quad \hat{u}^2(x^2,E^2) = (x^2 - h^2(E^2))^{1/2} \ . \tag{6.35}$$

This is how they think of their payoffs from 'fishing' in cardinally meaningful units. Of course, the $K^{\times}$ equilibrium of this game, in the economy $(\hat{u}^1,\hat{u}^2,G,X^{\Pr})$ is identical to the Kantian equilibrium with the utility representation $u^i(x^i,E^i) = x^i - h^i(E^i)$. Now the monotonic transformations which relate the second (obviously quasi-linear) representation to the representation of (6.35) are $f(u) = u^3$ and $g(u) = u^2$. By Proposition 6.5 (or equation (6.34)), the extended preferences that each must maximize in the game whose Nash equilibrium will be the Kantian equilibrium of the original game are $\hat{V}(\hat{u}^1,\hat{u}^2) = (\hat{u}^1)^{1/3} + (\hat{u}^2)^{1/2}$. But this is not a natural construction. The 'nice' formula $V(u^1,u^2) = u^1 + u^2$ only applies if the representation of the quasi-linear preferences is given by $u^i = x^i - h(E^i)$. So it is *false* to say that in a quasi-linear economy, it suffices to rationalize the Kantian equilibrium as a Nash equilibrium that each player *maximize total welfare.* This formulation only works for a *particular* representation of players' preference orders by utility functions.

Even in the case of quasi-linear *preferences* (i.e., preferences that admit a quasi-linear utility representation), the extended preferences that players would have to be using are complex. Let us state this in terms of laboratory experiments. Suppose the experimenter poses a prisoners' dilemma game to the players, whose payoffs are given by the matrix:

	Cooperate	Defect
Cooperate	(2,2)	(0,3)
Defect	(3,0)	(1,1)

(6.36)

According to Proposition 6.2, the simple Kantian equilibrium of this game is the same as the Nash equilibrium of the game where each player maximizes the *sum* of the payoffs $P^1 + P^2$. But now suppose the experimenter proposes the payoff matrix:

(6.37)

	Cooperate	Defect
Cooperate	(2,4)	(0,4.5)
Defect	(3,3)	(1,3.5)

Table (6.37) is, in fact, the same prisoners' dilemma game as table (6.36): I have simply transformed the von Neumann-Morgenstern utilities of the second player by the positive affine transformation $\frac{1}{2}u + 3$. The Kantian equilibria of the two games are identical. But in order to produce the Kantian equilibrium of the second game as a Nash equilibrium of a game with extended payoff functions, the players must both maximize $P^1 + 2(P^2 - 3)$. While maximizing the sum of payoffs might be natural, maximizing the latter function is not.

To reiterate, I take the results of this chapter to support the case that Kantian optimization is a fundamentally different protocol from Nash optimization[5]. Formally, we can explain cooperation as the result of Nash reasoning with preferences defined on the entire allocation. In some cases, the utility function that players must adopt on the entire allocation to rationalize Kantian equilibria in this way seems natural. But in general, the extended utility function is not natural. For example, the extended preferences that must be used for the Cobb-Douglas production economy of Proposition 6.4 cannot be represented by any function of the self-regarding utilities in the original game that can be written in closed form. Even for the simple case of quasi-linear

[5] It is true that I have investigated here only certain *kinds* of extended utility functions – ones whose arguments are the utilities of the players in the game. One might extend this study to more general extended preferences, which depend one the whole allocation of efforts and consumptions, but not necessarily through the utility functions of the players.

preferences, producing a simple extended utility function (which 'maximizes total utility') is only achieved with a particular choice of the utility representation of the original self-regarding preferences. The same is true with regard to rationalizing the prisoners' dilemma.

While Nash rationalization of cooperation is mathematically possible, to make the story credible, one would have to explain how players move from their self-regarding preferences to the extended preferences that are required. Behavioral economists have limited themselves to situations and experiments, almost ubiquitously, in which the notion of the fair allocation is obvious, or in which players have identical cardinal utility functions. In the latter case, the Nash equilibrium in which each 'maximizes total payoff' will coincide with the Kantian equilibrium of the game with self-regarding preferences. But this simple 'altruistic' preference order on the entire allocation only works for players with symmetric preferences represented by *particular* cardinal utility functions (or payoff functions), and does not work at all when players have different self-regarding preferences.

In contrast, the Kantian explanation seems much simpler. It enables cooperation where it is not obvious *what* the fair solution is (e.g., the production economies), or what the Pareto efficient solution is. Of course, to complete the story, one has to explain how people learn to optimize in the Kantian manner. When players are identical, simple Kantian optimization will work – and in this case, there is not much explaining to do. How people might learn to optimize in the multiplicative or additive way, for games with heterogeneous players, is harder to explain.

Words: 1955

Chapter 7. Existence and dynamics of Kantian equilibrium

All the results in chapters 4-6 are of the form: If a Kantian equilibrium exists, it has such-and-such properties. In this chapter, I prove that Kantian equilibria exist for a large domain of production economies, of the kind studied in chapters 4-6. Recall that a *proportional solution* for an economy $e = (u^1,...,u^n,G)$ is a Pareto efficient allocation in which consumption is proportional to efficiency units of effort expended. In Proposition 3.2, we showed that an economy's proportional solutions are precisely its strictly positive $K^\times$ equilibria.

The proportional solution was first defined and studied in Roemer and Silvestre (1993). There, a proof of the existence of proportional solutions was given for economies that include ones like *e,* but also much more complicated economies (with many goods and many kinds of labor). Here, I provide a simpler existence proof for a domain of economies like *e*. The proof, unlike that in the 1993 paper, exploits the fact that proportional solutions are precisely the positive $K^\times$ equilibria of the economy that allocates output in proportion to efficiency units of effort expended .

7.1 Existence of strictly positive $K^\times$ equilibria for production economies

Consider the following condition on utility functions $u(x,E)$:

Condition A. A utility function *u* satisfies Condition *A* if and only if

$\lim_{\varepsilon\to 0}\lim_{x\to 0}-\frac{u_2(x,\varepsilon)}{u_1(x,\varepsilon)} \to 0$.

Examples. Let $u(x,E) = x^a(1-E)^{1-a}$. Then $-\frac{u_2(x,\varepsilon)}{u_1(x,\varepsilon)} = \frac{1-a}{a}\frac{x}{1-\varepsilon}$, and Condition *A* holds. Let $u(x,e) = x - h(E)$ where $h'(0) = 0$. Then $-\frac{u_2(x,\varepsilon)}{u_1(x,\varepsilon)} = h'(\varepsilon)$ and Condition *A* holds. Let $u(x,E) = (ax^r + (1-a))(1-E)^r)^{1/r}$. Now, $-\frac{u_2(x,\varepsilon)}{u_1(x,\varepsilon)} = \frac{1-a}{a}(\frac{1-\varepsilon}{x})^{r-1}$ and Condition *A* holds for $-\infty < r \le 1$, which is to say, for all concave CES utility functions.

Define the domain of utility functions as:

$\mathbf{U} = \{u : \Re_+ \times [0,M] \to \Re,\ \text{some } M > 0, u \text{ differentiable \& concave, Condition } A \text{ holds}\}$.

Define the domain of production functions as:

$$\mathbf{G} = \{G : \Re_+ \to \Re_+; G \text{ increasing, differentiable, concave; } G' > 0\} \text{ .}$$

Define the domain of economies as $\mathbf{D} = \{(u^1,...,u^n,G) \mid u^i \in \mathbf{U}, G \in \mathbf{G}\}$.

Proposition 7.1 *For all economies* $e \in \mathbf{D}$ *, a strictly positive* K^X *equilibrium exists.*

Recall that a $K^\times$ *equilibrium* is a feasible allocation $(x^i,E^i)_{i=1,...,n}$ such that:

$$(\forall i) \quad x^i = \frac{E^i}{E^S} G(E^S) \text{ ,}$$

and for all i, $1 \in \arg\max_{0 \le \rho \le \frac{M}{E^i}} u^i(\frac{E^i}{E^S} G(\rho E^S), \rho E^i)$. (If $E^i = 0$, the domain of the argmax function is $0 \le \rho < \infty$.)

For any $\varepsilon > 0$, define the rectangle $R^\varepsilon = [\varepsilon, M]^n \subset \Re^n_{++}$. First, we prove a lemma:

Lemma 7.2 *For any* $G \in \mathbf{G}$, $\lim_{\varepsilon \to 0} \max_{\mathbf{E} \in R^\varepsilon} \frac{E^i}{E^S} G(\varepsilon \frac{E^S}{E^i}) = 0.$

Proof:

1. For any $\varepsilon > 0$, we must evaluate the solution of

$$\begin{aligned} &\max_Q \frac{1}{Q} G(\varepsilon Q) \\ &s.t. \\ &\frac{(n-1)\varepsilon + M}{M} \le Q \le \frac{(n-1)M + \varepsilon}{\varepsilon} \end{aligned} \tag{7.1}$$

where $Q = \frac{E^S}{E^i}$

There are three possibilities for the solution (a) $Q = 1 + \frac{(n-1)\varepsilon}{M}$; (b) $Q = 1 + \frac{(n-1)M}{\varepsilon}$, or (c) $Q \in (1 + \frac{(n-1)\varepsilon}{M}, 1 + \frac{(n-1)M}{\varepsilon})$.

2. Case (a). At $Q = 1 + \frac{(n-1)\varepsilon}{M}$, the maximand is

$$\frac{M}{M + (n-1)\varepsilon} G(\varepsilon + \frac{(n-1)}{M}\varepsilon^2) \to G(0) = 0 .$$

3. Case (b). In this case, the maximand is $\frac{\varepsilon}{\varepsilon + (n-1)M} G(\varepsilon + (n-1)M) \to 0$.

4. Case (c). In this case, we compute the first-order condition w.r.t. *Q,* for the maximization problem (7.1). This condition is:

$$\frac{G(\varepsilon Q^*)}{Q^*} = \varepsilon G'(\varepsilon Q^*) \tag{7.2}$$

at the solution Q^* . If εQ^* does not approach zero with ε , then the r.h.s. of (7.2) does approach zero, as we wish to show. If Q^* does approach zero, it cannot approach faster than $1 + \frac{(n-1)\varepsilon}{M}$, and so $\varepsilon G'(\varepsilon Q) \le \varepsilon G'(\varepsilon + \frac{(n-1)\varepsilon^2}{M}) \to \varepsilon G'(\varepsilon) \le G(\varepsilon) \to 0$. This proves the lemma. ■

Proof of Proposition 7.1:

1. Given $(\mathbf{u}, G) \in \mathbf{D}$. Consider the *n*-rectangle $R^\varepsilon = [\varepsilon, M]^n$, for some $0 < \varepsilon < M$. Define the individual best-reply correspondences on a domain R^ε as follows:

$$B^i(E^1, ..., E^n) = \{ rE^i \mid r \in \arg\max_{0 \le \rho \le \frac{M}{E^i}} u^i(\frac{E^i}{E^S} G(\rho E^S), \rho E^i) \} .^1 \tag{7.3}$$

[1] Notice the difference between these best-reply correspondences, which are defined on the entire effort vector, and the Nash-best-reply correspondences, which are defined on the on the vectors E^{-i} . The difference is illustrated graphically in figure 4.1.

Define $\mathbf{B} = (B^1,...,B^n)$, a mapping whose domain is R^ε and whose range is $\Re^n_+$. The mapping **B** is convex-valued because $u^i(\frac{E^i}{E^S}G(\rho E^S),\rho E^i)$ is a concave function of ρ. It is upper hemi-continuous by the Berge maximum theorem. We must show that **B** maps R^ε into itself, for some sufficiently small $\varepsilon > 0$. That is:

for some $\varepsilon > 0$ and any $\mathbf{E} = (E^1,...,E^n) \in R^\varepsilon$, $B^i(\mathbf{E}) \geq \varepsilon$ for all *i*.

The condition that guarantees the required inequality is:

$$(\exists \varepsilon > 0)(\forall i)(\forall \mathbf{E} \in R^\varepsilon)\left(\left.\frac{d}{dr}\right|_{r=\frac{\varepsilon}{E^i}} u^i(\frac{E^i}{E^S}G(rE^S),rE^i) \geq 0\right). \qquad (7.4)$$

For (7.4) guarantees that that scale factor *r* that maximizes *i*'s utility is at least $\frac{\varepsilon}{E^i}$ and hence $B^i(E^i,E^{N\setminus i}) \geq \frac{\varepsilon}{E^i}E^i = \varepsilon$. Condition (7.4) expands to:

$$(\exists \varepsilon > 0)(\forall i)(\forall \mathbf{E} \in R^\varepsilon)G'(\frac{\varepsilon}{E^i}E^S) \geq -\frac{u^i_2}{u^i_1}(\frac{E^i}{E^S}G(\frac{\varepsilon}{E^i}E^S),\varepsilon). \qquad (7.5)$$

The argument of G' is bounded above as ε approaches zero, so the left-hand side of this inequality is bounded away from zero, because $G \in \mathbf{G}$. It therefore suffices to show that the right-hand side approaches zero as ε becomes small.

Now Lemma 7.2 tells us that the argument $\frac{E^i}{E^S}G(\frac{\varepsilon}{E^i}E^S)$ of the marginal rate of substitution on the right-hand side of (7.5) approaches zero, and therefore, by Condition *A*, the marginal rate of substitution approaches zero.

This concludes the demonstration that for small enough ε, **B** maps R^ε into itself.

2. Therefore, all the assumptions of Kakutani's fixed point theorem hold, and so a fixed point of **B** exists on some domain R^ε. But a fixed point on this domain is a strictly positive K^X equilibrium, which concludes the proof. ■

7.2 Existence of K^β equilibria for $0 < \beta \leq \infty$

Recall (chapter 4) the infinite family of efficient Kantian pairs (X_β,φ_β), for $0\le\beta\le\infty$. Proposition 7.1 proves the existence of positive Kantian equilibria for $\beta=0$, which is to say, Pareto efficient Kantian equilibria. The proof for $\beta>0$ is simpler, because of Proposition 4.1: that is, any K^β Kantian equilibrium for $\beta>0$ is Pareto efficient. (A K^β equilibrium is one for the efficient Kantian pair (X_β,φ_β).) We do not need an analogue to Lemma 7.2.

Let $\tilde{\mathbf{U}}=\{u:\Re_+\times[0,M]\to\Re,\ \text{some } M>0, u \text{ differentiable \& concave}\}$; let $\tilde{\mathbf{G}}=\{G:\Re_+\to\Re_+, G \text{ concave \& differentiable}\}$; let $\tilde{\mathbf{D}}=\{e=(u^1,...,u^n,G), u^i\in\tilde{\mathbf{U}}, G\in\tilde{\mathbf{G}}\}$.

Proposition 7.3 *For any $\beta>0$, on the domain $\tilde{\mathbf{D}}$, a Kantian equilibria with respect to the allocation rule and Kantian variation (X_β,φ_β) exists.*

Proof:

1. Consider $0<\beta<\infty$. The domain of effort vectors is the convex, compact set $R^0=[0,M]^n$. For any vector in the domain, define the best-reply correspondences:

$$B^i(E^1,...,E^n)=\{rE^i+(r-1)\beta \mid r\in \underset{\rho\in[\frac{\beta}{E^i+\beta},\frac{M+\beta}{E^i+\beta}]}{\arg\max}\ u^i\left(\frac{E^i+\beta}{E^S+n\beta}G(\rho E^S+n(\rho-1)\beta),\rho E^i+(\rho-1)\beta\right) \quad (7.6)$$

The domain restriction on ρ in the maximization guarantees that $B^i(E^1,...,E^n)\subseteq R^0$. Define the vector-valued best-reply correspondence by $\mathbf{B}=(B^1,...,B^n)$. $\mathbf{B}$ is convex-valued, because the argument in the maximization in (7.6) is a concave function of ρ. It is upper hemi-continuous by the maximum theorem. Hence all the assumptions of the Kakutani fixed point theorem hold, and so a fixed point of $\mathbf{B}$ exists. But such a fixed point is a K^β equilibrium.

2. For $\beta=\infty$ (that is, K^+ equilibrium), a separate argument is required. It is obvious how to define the best-reply correspondence, and the same approach works. ■

7.3 Is there an allocation rule which Nash equilibrium implements efficiently on the domain $\tilde{\mathbf{D}}$?

What Propositions 7.1 and 7.3 show is that for every Kantian variation K^{β} , for $0 \le \beta \le \infty$ there is (at least) one allocation rule that optimization according to the K^{β} protocol implements efficiently on a large domain of production economies. We may ask: Is there *any* allocation rule that the Nash protocol implements efficiently on the domain $\tilde{\mathbf{D}}$? The answer is negative.

Proposition 7.4

A. *There is no allocation rule that is efficiently implementable in Nash equilibrium on the domain* $\tilde{\mathbf{D}}$.

B. *On continuum economies, Walrasian rules (with no taxation) are efficiently Nash implementable*.

The Walrasian allocation rules are defined by the following way of sharing output. The share allocated to player *i* is equal to:

$$\text{for all } i, \quad \theta^{i,Wa}(E^1,...,E^n) = \frac{G'(E^S)}{G(E^S)}E^i + \sigma^i(1 - \frac{G'(E^S)E^S}{G(E^S)}), \tag{7.7}$$

where the profit shares $(\sigma^1,...,\sigma^n)$ are fixed non-negative numbers summing to one. Equations (7.7) state that the output received by player i (that is, $\theta^{i,Wa}G(E^S)$) is equal to the sum of her labor income (her effort supply times the marginal productivity of effort), and her share of profits.

Proof of Proposition 7.4:

1. In this proof, we specify an allocation rule by the shares of output that are allocated to each player, as a function of the effort vector **E**. An interior allocation **E** is a Nash equilibrium on the domain of economies for the allocation rule θ if and only if

$$\forall j \quad u_1^j \cdot (\frac{\partial \theta^j(\mathbf{E})}{\partial E^j}G(E^S) + \theta^j(\mathbf{E})G'(E^S)) + u_2^j = 0 \ . \tag{7.8}$$

Therefore θ is efficiently Nash-implementable if and only if:

$$\forall j \quad 1 = \theta^j(\mathbf{E}) + \frac{G(E^S)}{G'(E^S)} \frac{\partial \theta^j(\mathbf{E})}{\partial E^j} \ . \tag{7.9}$$

2. Indeed, (7.9) must hold for the entire positive orthant $\Re^n_{++}$, for given any positive vector **E**, we can construct *n* concave utility functions such that (7.8) holds at **E**.

3. For fixed **E**, define $\psi^j(x) = \theta^j(E^1, E^2, ..., E^{j-1}, x, E^{j+1}, ..., E^n)$ and $\mu^j(x) = G(x + E^S - E^j)$. Then (7.9) gives us the differential equation:

$$1 = \psi^j(x) + \frac{\mu^j(x)}{(\mu^j)'(x)} (\psi^j)'(x) \ , \tag{7.10}$$

which must hold on $\Re_{++}$.

4. But (7.10) implies that

$$\frac{(\psi^j)'(x)}{1 - \psi^j(x)} = \frac{(\mu^j)'(x)}{\mu^j(x)} \ , \tag{7.11}$$

which implies that $\mu^j(x)(1 - \psi^j(x)) = k^j$ and therefore $\psi^j(x) = 1 - \frac{k^j(E^{-j})}{\mu^j(x)}$ where the constant k^j may depend on the ray $(E^1, .., E^{j-1}, x, E^{j+1}, .., E^n)$ on which ψ^j is defined – that is, upon E^{-j} .

5. In turn, this last equation says that on the ray $(E^1, ..., E^{j-1}, x, E^{j+1}, ..., E^n)$ we have:

$$\theta^j(E^1, ..., E^{j-1}, x, E^{j+1}, ..., E^n) G(x + E^S - E^j) = G(x + E^S - E^j) - k^j(E^{-j}) \ , \tag{7.12}$$

which says that 'every agent receives his entire marginal product' on this space. To be precise:

$$(\forall x, y > 0)$$
$$(\theta^j(E^1,...,E^{j-1},x,E^{j+1},...,E^n)G(x+E^S-E^j)-$$
$$\theta^j(E^1,...,E^{j-1},y,E^{j+1},...,E^n)G(y+E^S-E^j)=$$
$$G(x+E^S-E_j)-G(y+E^S-E_j))\ .$$

(7.13)

Now let $y=0$ and $x=E^j$ and let $z^j=\theta^j(E^1,...,E^{j-1},0,E^{j+1},...,E^n)G(E^S-E^j)$. Then (7.14) says that:

$$(\forall j)(\theta^j(E)G(E^S)-z^j=G(E^S)-G(E^S-E^j))\ . \qquad (7.15)$$

.

6. Adding up the equations in (7.15) over *j*, and using the fact that $z^j \geq 0$, we have:

$$G(E^S) \geq nG(E^S)-\sum G(E^S-E^j)$$

or :
$$G(E^S) \leq \frac{1}{n-1}\sum G(E^S-E^j)\ . \qquad (7.16)$$

7. Now note that $\frac{1}{n-1}\sum(E^S-E_j)=E^S$. Therefore (7.16) can be written:

$$G(\frac{1}{n-1}\sum(E^S-E_j)) \leq \frac{1}{n-1}\sum G(E^S-E_j)\ , \qquad (7.17)$$

which is impossible for any strictly concave *G*. This proves part A of the theorem.
8. The proof of part B is well-known: for part B just says that Nash behavior, taking prices as given, at the Walrasian allocation rule, induces Pareto efficiency. ■

The key point, in part A of Proposition 7.4, is that in a finite economy, an agent cannot ignore the effect of his labor supply on the marginal productivity of labor. It is only in an economy with an infinite number of agents that the wage is not affected by individual labor choices.

Proposition 7.4 is another way of stating the benefits of cooperation. Of course, cooperation only works on a 'small' domain of allocation rules for production economies:

the rules that allocate part of the product according to equal division, and part according to proportional division.[2]

7.4 Dynamics

There is a convenient dynamic process that, for well-behaved games, will converge to a Nash equilibrium of the game from an arbitrary initial strategy vector. It is to iterate the best-reply correspondence. If the payoff functions are well-behaved, 'iterated best replies' converges to a Nash equilibrium of the game. We can use the same procedure for Kantian equilibrium: 'iterated best replies' converges to a Kantian equilibrium, if the game is well-behaved. The purpose of this section – to demonstrate this -- is again to emphasize the formal similarity between Nash and Kantian equilibrium.

We will study a special case: there are two players, they each have quasi-linear preferences $u^i(x,E) = x - c^i(E)$, for i =1,2, where c^i are strictly convex, increasing, differentiable functions. We will work with the K^+ protocol and the equal-division rule. We are given an economy (u^1,u^2,G) with G concave. Define the best-reply function for effort vectors in $R^0 = [0,M]^2$:

$$\mathbf{B} = (B^1,B^2), \text{ where } B^i(E^1,E^2) = E^i + r^i(\mathbf{E}) \qquad (7.18)$$

and $r^i(\mathbf{E}) = \underset{-E^i \le r \le M - E^i}{\arg\max}\, u^i(\frac{G(E^S + 2r)}{2}, E^i + r)$. For the utility functions specified, the argmax in (7.18) is unique, and so **B** is a single-valued. Note that a fixed point of **B** is an additive Kantian equilibrium for the equal-division rule, since if $\mathbf{B}(E^1,E^2) = (E^1,E^2)$ then "$r = 0$" is the argmax for both players.

Proposition 7.5 *The mapping* $\mathbf{B}: \Re^2_+ \to \Re^2_+$ *is a contraction mapping, and hence possesses a unique fixed point.*[3]

---

[2] As I indicated earlier, there are other rules that I do not describe here that can be efficiently Kantian-implemented, that are in 'neighborhoods' of these rules.

[3] It is a well-known mathematical fact that a contraction mapping possesses a unique fixed point, and that iterated application of the mapping from any initial point converges to the fixed point.

It immediately follows that iterated application of $\mathbf{B}$ starting from any initial vector of efforts will converge to its fixed point, which is an additive Kantian equilibrium (indeed, the unique such equilibrium for the economy specified).

The proof of Proposition 7.5 uses a well-known mathematical fact:

Lemma *Let* $\|\ \|$ *be a norm on* $\Re^n$ *and let* $[\![A]\!]$ *be the associated sup norm on mappings* $A:\Re^n \to \Re^n$, *defined by* $[\![A]\!] = \sup_{\|x\|=1} \|A(x)\|$. *Let* $J(A)$ *be the Jacobian matrix of A. If* $[\![J(A)]\!] < 1$, *then A is a contraction mapping.*

Proof of Proposition 7.5:

1. The Jacobian of the mapping $\mathbf{B}$ is $\begin{pmatrix} 1+r_1^1 & r_2^1 \\ r_1^2 & 1+r_2^2 \end{pmatrix}$, where

$r_i^j(E^1,E^2) = \dfrac{\partial r^j}{\partial E^i}(E^1,E^2)$, assuming that these derivatives exist. Thus, the lemma requires that we show the norm of this matrix is less than unity. We take $\|\ \|$ to be the Euclidean norm on $\mathbb{R}^2$. We must show that:

$$\|E\| = 1 \Rightarrow \left\| \begin{pmatrix} 1+r_1^1(E) & r_2^1(E) \\ r_1^2(E) & 1+r_2^2(E) \end{pmatrix} \begin{pmatrix} E^1 \\ E^2 \end{pmatrix} \right\| < 1 \ . \tag{7.19}$$

2. By differentiability of c^j, the function r^j is defined by the following first-order condition:

$$G'(E^S + 2r^j(\mathbf{E})) = (c^j)'(E^j + r^j(\mathbf{E})) \ , \tag{7.20}$$

which has a unique solution. By the implicit function theorem, the derivatives of $r^j(\cdot)$ are given by:

$$G''(y^j)(1+2r_i^j(\mathbf{E})) = (c^j)''(x^j)(\delta_i^j + r_i^j(\mathbf{E})) \,,$$

where $y^j = G(E^S + 2r^j(\mathbf{E}))$, $x^j = E^j + r^j(E)$ and $\delta_i^j = \begin{cases} 1, \text{ if } i = j \\ 0, \text{ if } i \neq j \end{cases}$; or

$$r_i^j(E) = \frac{\delta_i^j (c^j)''(x^j) - G''(y^j)}{2G''(y^j) - (c^j)''(x^j)} \ . \tag{7.21}$$

3. It follows from (7.19) that the Jacobian of $\mathbf{B}$ is given by:

$$\begin{pmatrix} \dfrac{G''(y^1)}{2G''(y^1) - (c^1)''(x^1)} & \dfrac{-G''(y^1)}{2G''(y^1) - (c^1)''(x^1)} \\ \dfrac{-G''(y^2)}{2G''(y^2) - (c^2)''(x^2)} & \dfrac{G''(y^2)}{2G''(y^2) - (c^2)''(x^2)} \end{pmatrix}$$

and so, from step 1, we need only show that:

$$(Q^1(E^1 - E^2))^2 + (Q^2(E^1 - E^2))^2 < 1 \ , \tag{7.22}$$

where $\left\| (E^1, E^2) \right\| = 1$ and $Q^j = \dfrac{G''(y^j)}{2G''(y^j) - (c^j)''(x^j)}$. Note that $\left| Q^j \right| < \dfrac{1}{2}$. Inequality (7.22) reduces to showing that $\dfrac{1}{2}(1 - 2E^1E^2) < 1$, which is obviously true, proving the proposition. ■

Word count: 3498

Chapter 8. Evolutionary considerations

In this chapter, we examine several simple environments in which Kantian and Nash players meet each other repeatedly and play a game. The question is whether Kantian players can resist invasion by Nash players.

We assume there is a population, fraction v of whom are Kantian optimizers, and fraction $1-v$ of whom are Nash optimizers (henceforth, Nashers). At each date, individuals from this population are randomly paired and play a game. The fitness of each group is a strictly monotone increasing function of the average payoff of the members of that group. The population is stable when the fitness of both Kantian and Nash players is the same. If the fitness of Nashers is greater than the fitness of Kantians for all v, then Nashers drive Kantians to extinction, and conversely. We consider two games: the random dictator game, and the general 2×2 symmetric game with mixed strategies.

It is assumed that when two agents are matched to play a game, they cannot recognize each other's type. If Kantians could recognize their opponent's type, they could simply play Kantian when matched with Kantians and play Nash when matched with Nash players, and their average payoff would be greater than the average payoff of Nash players in the games we study. Therefore Kantians would drive Nashers to extinction. The problem of invasion is therefore only interesting when Kantians cannot recognize the type of their opponent.

8.1 Random dictator game

We assume that all Kantians possess a von Neumann-Morgenstern utility function u over money payoffs which is risk averse, and we normalize u by:

$$u(0)=0,\quad u(\frac{1}{2})=\rho,\quad u(1)=1 \text{ where } \rho>\frac{1}{2}\ . \tag{8.1}$$

In the random dictator game, one of the two players is chosen to be the dictator randomly by Nature; the dictator divides one unit of resource between himself and the other player. We have shown in chapter 1 that the simple Kantian equilibrium for two risk averse Kantian players is to split the resource equally between them. I propose that, in a situation where one does not know the type of one's opponent, Kantian players adopt this