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

strategy: with some probability π, play the Kantian strategy if chosen dictator, and with probability $1-\pi$, play the Nash strategy. The optimization problem for the Kantian is to choose π, knowing *v*. We then calculate the payoffs of the two player types.

Suppose, then, all Kantian players will choose to play Kantian with probability π. We suppose that fitness and fertility is a an increasing function of one's payoff. Without loss of generality, we measure fitness by *u*. If a Kantian gives $1/2$ to his opponent, if chosen to be dictator, with probability π and 0 to his opponent with probability $1-\pi$, then she faces the following events:

- with probability $\pi/2$ she is dictator and keeps ½
- with prob $(1-\pi)/2$ she is dictator and keeps 1
- with prob $v\pi/2$ she faces a Kantian who is dictator, and she gets ½
- with prob $v(1-\pi)/2$ she faces a Kantian who is dictator and she gets 0
- with prob $(1-v)/2$ she faces a Nash player who is dictator and she gets 0.

It follows that the average fertility of Kantians is:

$$\frac{\pi}{2}u(\frac{1}{2})+\frac{1-\pi}{2}u(1)+v\frac{\pi}{2}u(\frac{1}{2})+u(0)(\frac{v(1-\pi)+1-v}{2})= \frac{\pi}{2}(1+v)u(1/2)+\frac{1-\pi}{2}u(1) \tag{8.2}$$

For the Nash player:

- with prob ½ he is chosen dictator and keeps 1
- with prob $(1-v)/2$ he meets a Nash player who is chosen dictator and he gets 0
- with prob $v\pi/2$ his Kantian opponent, who is dictator, gives him ½
- with prob $v(1-\pi)/2$ his Kantian opponent, who is dictator, gives him 0.

Therefore the expected fertility of Nash players is:

$$u(1)/2+\frac{v\pi}{2}u(\frac{1}{2})+u(0)(\frac{v(1-\pi)+1-v}{2})=u(1)/2+\frac{v\pi}{2}u(\frac{1}{2})\ . \tag{8.3}$$

Thus fertilities are equal if and only if $\frac{\pi}{2}(1+v)u(\frac{1}{2})+\frac{1-\pi}{2}u(1)=\frac{u(1)}{2}+\frac{v\pi}{2}u(\frac{1}{2})$

or $\frac{\pi}{2}u(\frac{1}{2})=\frac{\pi}{2}u(1)$ which happens if and only if $\pi=0$. It therefore follows that there can be no stable equilibrium with both types present unless Kantian players behave exactly like Nash players – that is, they always play Nash.

Now let's look at Kantian's optimal choice of π . Her expected utility is

$$\frac{\pi}{2}u(\frac{1}{2})+\frac{1-\pi}{2}u(1)+v\frac{\pi}{2}u(\frac{1}{2})=$$
$$\frac{\pi}{2}\rho+\frac{1-\pi}{2}+v\frac{\pi}{2}\rho=\frac{\pi}{2}(\rho(1+v)-1)+\frac{1}{2}.$$

whose derivative with respect to ρ is $\frac{\rho(1+\nu)-1}{2}$.

So her choice is:

$$\pi=\begin{cases}1, & \text{if } v>\frac{1}{\rho}-1\\ [0,1], & \text{if } v=\frac{1}{\rho}-1\\ 0, & \text{if } v<\frac{1}{\rho}-1\end{cases} \qquad (8.4)$$

Suppose we begin with a population consisting entirely of Kantians -- $v=1$. Then Nash players appear. Kantians will continue playing the pure Kantian strategy until *v* falls to $v^*=\frac{1}{\rho}-1$. At this point, according to (8.4), Kantians should randomize completely. The average fertility of Nash players continues to be higher than that of Kantian players. In this game with a continuum of players, the equilibria only exist when $v<\frac{1}{\rho}-1$, because otherwise, a non-negligible fraction of Kantians will be playing Kantian with positive probability, and their payoff will be less than the Nash payoff. Although an equilibrium in the dynamic process postulated here does not exist (it would

have to be arbitrarily close to but less than p^*), we can at least assert that we will not see a stable population where Kantian play is observed. To survive, all Kantians have to play Nash strategies almost surely.

8.2 2×2 symmetric games

We study games whose payoff matrices are given by

	X	Y
X	(1,1)	(a,b)
Y	(b,a)	(0,0)

where $(a,b)\in\Re^2$, a two-dimensional set of games. We assume the games are *generic,* which is defined to mean that $0\neq a\neq 1, 0\neq b\neq 1, 1\neq a+b\neq 2$. Define the following subsets of the (a,b) plane:

$$N_1=\{b<1\}$$
$$N_2=\{a<0\}$$
$$N_3=\{a>0, b>1, a+b>1\}$$
$$N_4=\{a<0, b<1, a+b<1\}$$
$$K_1=\{a+b>2\}$$
$$K_2=\{a+b<2\}$$

We will study Nash and simple Kantian equilibria of these games, when players play mixed strategies over *X* and *Y*. The strategy 'play *X* with probability *p* and *Y* with probability $1-p$ ' will be denoted *p*. Define the mixed strategies

$$p_1^*=\frac{a+b}{2(a+b-1)},\quad q_1^*=\frac{a}{a+b-1}\ , \tag{8.5}$$

in the case where the two probabilities so defined are in the interval $(0,1)$.

The structure of the dynamic economy is as was described in the previous section. There is a frequency of Kantians *v* in the economy. At each date, agents are paired off and play the (a,b) game, for fixed (a,b) . Denote the simple Kantian equilibrium of the game in question by p^* and the (or one of the) Nash equilibrium(a) of the game by q^* . Nashers always play the mixed strategy q^* . But Kantians, who will in general be

exploited by Nashers, play a compound strategy: with some probability π a Kantian plays p^* and with probability $1-\pi$ she plays q^*. Given *v*, all Kantians choose π to maximize their expected payoff. The question is whether, for a given (a,b), there is a stable mixed population of types, or whether one type drives the other to extinction.

We first characterize the Nash and simple Kantian equilibria for these games.

Lemma 8.1[1] *In the* (a,b) *game, the Nash equilibrium* (NE) *is:*

$$(X,X) \Leftrightarrow N_1 ,$$
$$(Y,Y) \Leftrightarrow N_2$$
$$(q_1^*, q_1^*) \Leftrightarrow N_3 \cup N_4$$

and the simple Kantian equilibrium is:

$$(p_1^*, p_1^*) \Leftrightarrow K_1$$
$$(X,X) \Leftrightarrow K_2 \quad .$$

Proof:

1. Let the row player play *p* and the column player *q*. Then payoff to the row player is:

$$V(p,q) = pq + p(1-q)a + (1-p)qb = \\ p(q + (1-q)a - qb) + qb \qquad (8.6)$$

which is linear in *p*. Hence the best (Nash) response of the row player, *p,* is given by:

$$p = 1 \text{ if } q + (1-q)a > qb$$
$$p = 0 \text{ if } q + (1-q)a < qb \quad .$$

It follows that (*X,X*) is a NE if and only if N_1 and (*Y,Y*) is a NE iff N_2 .

If $q + (1-q)a = qb$ -- which is to say, $q = q_1^*$ -- then any *p* is a best response. Thus (q_1^*, q_1^*) is a NE exactly when $q_1^* \in [0,1]$, which means $N_3 \cup N_4$. Thus, the Nash equilibria are characterized. Clearly, there is a region with multiple NE.

2. To compute the SKE, we maximize the symmetric payoff

$$V^K(p,p) = p^2 + p(1-p)a + (1-p)pb = \\ p^2(1-a-b) + p(a+b).$$

[1] The notation hereafter identifies a set N_i with the class of games such that $(a,b) \in N_i$, etc.

If $1 < a+b$, this is a concave function of p, and the first-order condition gives the maximum -- $p = \frac{a+b}{2(a+b-1)}$. Thus in this case, the solution is:

$$p = \begin{cases} p_1^*, \text{ if } K_1 \\ 1, \text{ if } 1 < a+b < 2 \end{cases} .$$

The second case is that $1 > a+b$. Then V^K is a convex function of p, and the solution is either $p = 0$ or $p = 1$. Check that $V^K(1,1) > V^K(0,0)$ so it is $p = 1$. It follows that $p = 1$ is the SKE when $1 < a+b < 2$ or $a+b < 1$, which is to say when K_2 . This concludes the characterization. ■

Definition A *pure coordination game* is one in which there are multiple pure-strategy Nash equilibria, and one equilibrium Pareto dominates the others.

Lemma 8.2 *The pure coordination* (a,b) *games are precisely those in* $N_1 \cap N_2$.

Proof:

From lemma 8.1, the games in $N_1 \cap N_2 = \{a < 0\} \cap \{b < 1\}$ are pure-coordination games, because they possess two pure-strategy Nash equilibria, (X,X) and (Y,Y) which are Pareto ranked. Are there pure coordination games whose Pareto-ranked equilibria are not the symmetric equilibria (X,X) and (Y,Y) ? The answer is no. One easily verifies, using equation (8.6), that (X,Y) is a Nash equilibrium iff $\{a < 0 \,\&\, b > 1\}$, and (Y,X) is a Nash equilibrium iff $\{a > 0 \,\&\, b < 1\}$. Therefore (X,X) and (X,Y) are Nash equilibria iff $\{a < 0 \,\&\, b > 1\}$, but in this region these equilibria are not Pareto ranked since neither payoff vector (1,1) nor (a,b) dominates the other. There is no region in which both (X,Y) and (Y,X) are both Nash equilibria. In other regions where multiple pure-strategy Nash equilibria exist, the equilbria are not Pareto ranked. ■

Definition. An (a,b) game is *supermodular* if $\frac{\partial^2 V(p,q)}{\partial p \, \partial q} > 0$: that is, iff $1 > a+b$.

The condition that defines supermodularity is often called strategic complementarity. One might say of these games that 'increased cooperation begets increased cooperation.' We have:

Lemma 8.3[2] *An* (a,b) *game is supermodular and possesses a mixed strategy Nash equilibrium if and only if it is in* $N_1 \cap N_2$.

Proof:

From lemma 8.1, a mixed-strategy Nash equilibrium exists iff $N_3 \cup N_4$. A game is supermodular iff $a+b<1$. The intersection of these two conditions is $\{a<0 \,\&\, b<1\} = N_1 \cap N_2$. ■

To study the behavior of Nash and Kantian players when they meet, we must examine each region $N_n \cap K_k$, for $1 \le n \le 4, 1 \le k \le 2$. The region $N_4 \cap K_1$ is empty. This leaves seven non-empty regions, which we name as follows, with their associated mixed and pure strategy (symmetric) equilibria for the two player types[3]:

Region number	**Region**	**NE**	**SKE**
I	$N_1 \cap K_1$	X	p_1^*
II	$N_2 \cap K_1$	Y	p_1^*
III	$N_3 \cap K_1$	q_1^*	p_1^*
IV	$N_1 \cap K_2$	X	1
V	$N_2 \cap K_2$	Y	1
VI	$N_3 \cap K_2$	q_1^*	1
VII	$N_4 \cap K_2 = N_4$	q_1^*	1

Table 8.1 Characterization of symmetric NE and SKE in 2×2 symmetric games

[2] I thank Burak Unveren for this result.

[3] Some of these regions intersect; in the intersection, there are multiple Nash equilibria. Thus a 'Region' is not simply a region of the (a,b) plane, but it's a planar region *and* a choice of Nash and Kantian equilibria in that region.

Denote the Nash equilibrium in a region by q^* and the simple Kantian equilibrium by p^*, where (p^*,q^*) takes on the values in Table 8.1, as a function of the region. (If the Nash equilibrium is (X,X) then $p^*=1$, etc.) As I have said, we suppose that, given a frequency *v* of Kantian players in the population, a Kantian will play a compound strategy $z(\pi)=\pi p^*+(1-\pi)q^*$, where π is chosen to maximize the Kantian's expected payoff, given *v*. There are four possible outcomes for a Kantian player in the game, whose payoffs are given by:

Strategy profile	Probability x Payoff to Kantian
X,X	$z(\pi)(vz(\pi)+(1-v)q^*)\cdot 1$
X,Y	$z(v(1-z)+(1-v)(1-q^*)a$
Y,X	$(1-z)(vz+(1-v)q^*)b$
Y,Y	0

Table 8.2 Expected payoffs to a Kantian player

We read this table as follows. What is the probability that the Kantian and her opponent both play *X*? She plays *X* with probability $z(\pi)$; her opponent will be a Kantian who plays *X* with probability $vz(\pi)$ and he will be a Nasher who plays *X* with probability $(1-v)q^*$. The conjunction of these events occurs with probability $z(\pi)(vz(\pi)+(1-v)q^*)$, and in this case the payoff to the Kantian is 1, giving the first row of the table. And so on. Therefore, we can define the expected payoff to the Kantian player, if she plays the compound strategy $z(\pi)$ and the frequency of Kantians is *v* by the sum of the elements in the second column in table 8.2:

$$\mathbf{V}^K(\pi,v)=z(vz+(1-v)q^*)+z(v(1-z)+(1-v)(1-q^*))a+ (1-z)(vz+(1-v)q^*)b. \tag{8.7}$$

In like manner, compute that the expected payoff to a Nasher is given by:

$$\mathbf{V}^N(\pi,v) = q^*(vz+(1-v)q^*)+q^*(v(1-z)+(1-v)(1-q^*))a+ \\ (1-q^*)(vz+(1-v)q^*)b. \tag{8.8}$$

The average payoff (i.e., fitness) of Kantians is greater than the average payoff of Nashers if and only if $\mathbf{V}^K(\pi,v) > \mathbf{V}^N(\pi,v)$, which occurs precisely when:

$$(z-q^*)(vz+(1-v)q^*)+(z-q^*)(v(1-z)+(1-v)(1-q^*))a+ \\ (q^*-z)(vz+(1-v)q^*)b>0. \tag{8.9}$$

Dividing this inequality by $z-q^*$ and simplifying, we can write the condition as consisting of two cases:

$$vz(1-a-b) > (1-v)q^*(a+b-1)-a, \text{ if } z>q^* , \tag{8.10}$$

and

$$vz(1-a-b) < (1-v)q^*(a+b-1)-a, \text{ if } z<q^* . \tag{8.11}$$

In principle, we must check whether the appropriate inequality holds for each region, when $z=z(\pi)$ is evaluated at the optimal value of π, defined by:

$$\pi^*(v) = \arg\max_{\pi} \mathbf{V}^K(\pi,v) . \tag{8.12}$$

A necessary condition for the fitness of Kantians to (weakly) exceed the fitness of Nashers is that:

$$(\exists \pi>0)(\mathbf{V}^K(\pi,v) \geq \mathbf{V}^N(\pi,v)) . \tag{8.13}$$

In other words, if (8.13) is false for all values of *v*, then Nashers surely drive Kantians to extinction, unless Kantians act just like Nashers and play the Nash strategy with probability one.

Proposition 8.4

A. *For games in Region* VII = $\{a<0\}\cap\{b<1\} = N_1 \cap N_2$, *Kantians drive Nashers to extinction. This is true whether Nashers play either of their symmetric-equilibrium strategies,* $q^* \in \{0, q_1^*\}$.

B. In all other regions, either Kantians and Nashers play identically, or Nashers drive Kantians to extinction.

C. Generically, there are no games in which Kantians and Nashers play different strategies and coexist at stable frequencies.

Proof:

We examine the games by Region.

Region I $(p^*,q^*)=(\frac{a+b}{2(a+b-1)},1)$.

Since $p^*<q^*$, $z(\pi)<q^*$ for all positive π . Since $a+b>2>1$ in this region, (8.11) holds iff $z>\frac{(1-v)(a+b-1)-a}{v(1-a-b)}$. There exists a positive value of π such that this inequality holds iff :

$$q^*=1>\frac{(1-v)(a+b-1)-a}{v(1-a-b)}. \tag{8.14}$$

(8.14) reduces to the inequality $b>1$, which is false in this region. Indeed, for all positive π, and all *v,* it follows that $\mathbf{V}^N(\pi,v)>\mathbf{V}^K(\pi,v)$, and so Nashers drive Kantians to extinction, unless Kantians imitate Nashers, in which case we see no Kantian behavior.

Region II $(p^*,q^*)=(\frac{a+b}{2(a+b-1)},0)$

Since $p^*>q^*$, $z(\pi)>q^*$ for all positive π , so we analyze (8.10), which reduces to:

$$z<\frac{a}{v(a+b-1)} . \tag{8.15}$$

There exists a positive π such that (8.15) is true iff $0<\frac{a}{a+b-1}$, which is false in this region, so Nashers drive Kantians to extinction.

Region III $(p^*,q^*)=(\frac{a+b}{2(a+b-1)},\frac{a}{a+b-1})$

There are two sub-regions:

Region IIIa $b>a$. Here, $p^*>q^*$, so $z(\pi)>q^*$ for all positive π . We analyze (8.10), which reduces to:

$$z<\frac{(1-v)q^*(a+b-1)-a}{v(1-a-b)} . \tag{8.16}$$

There is a positive value such that (8.16) is true iff:

$$q^*<\frac{(1-v)q^*(a+b-1)-a}{v(1-a-b)} , \tag{8.17}$$

which reduces to $a < a$, an impossibility. Moreover, $\mathbf{V}^K(\pi, v) = \mathbf{V}^N(\pi, v)$ iff $\pi = 0$. So Nashers drive Kantians to extinction.

Region IIIb, $b < a$. Here, $p^* < q^*$, so $z(\pi) < q^*$. We analyze condition (8.11), which reduces:

$$z > \frac{(1-v)q^*(a+b-1)-a}{v(1-a-b)} \quad . \tag{8.18}$$

There exists a positive π such that (8.18) is true iff

$$q^* > \frac{(1-v)q^*(a+b-1)-a}{v(1-a-b)} \quad , \tag{8.19}$$

which reduces to the inequality $a < a$. Again, Nashers drive Kantians to extinction.

Region IV. The games in this region are uninteresting because both types play the same strategy $p^* = q^* = 1$.

Region V. $(p^*, q^*) = (1,0)$

It follows that $z(\pi) > q^*$ for positive π , so we analyze condition (8.10). There are three sub-cases.

Region Va. $a < 0, a+b < 1$ and $b < 1$.

We compute $\pi^*(v)$. The coefficient of z^2 in $\mathbf{V}^K(\pi, v)$ is $v(1-a-b) > 0$, so $\pi^*(v)$ is either 0 or 1. It is 1 iff $v + (1-v)a > 0$, or $v > -\frac{a}{1-a}$. This inequality is true for $v = 1$, since $a < 0$. Substitute $\pi = 1$ (and therefore $z = 1$) into (8.10), and note that the inequality holds, since $b < 1$. Therefore, in this region, Kantians resist invasion by Nashers.

Region Vb. $a < 0, a+b < 1$ and $b > 1$.

In this case, (8.10) is true iff $vz(1-a-b) > -a$. Since this inequality fails at $v \in \{0,1\}$, for any *z,* it fails for all *v* for any *z.* Therefore Nashers drive Kantians to extinction.

Region Vc. $a < 0$ and $2 > a+b > 1$.

Inequality (8.10) holds iff $vz(1-a-b) > -a$, which is false, because the l.h.s. is negative or zero and the r.h.s. is positive. So Nashers drive Kantians to extinction.

Region VI. $(p^*,q^*)=(1,\frac{a}{a+b-1})$.

Since $p^*>q^*$, $z(\pi)>q^*$. We examine (8.10), which can be written:

$$z<\frac{-a}{1-a-b} \quad . \tag{8.20}$$

A positive value of π exists for which (8.20) is true iff $q^*=\frac{a}{a+b-1}<\frac{-a}{1-a-b}$. But this is false, and so Nashers drive Kantians to extinction.

Region VII. $(p^*,q^*)=(1,\frac{a}{a+b-1})$.

In this region, $z(\pi)>q^*$; we examine (8.10), which reduces to:

$$z>\frac{a}{a+b-1}=q^* \quad , \tag{8.21}$$

which is true for all $\pi>0$. Therefore (8.21) is true for $\pi=\pi^*(1)$, as long as $\pi^*(1)>0$. Compute that $\mathbf{V}^K(\pi,1)=z^2+z(1-z)(a+b)$, which is a convex function of z and hence of π . It is maximized at $\pi=1$ iff $1>\frac{-a^2}{a+b-1}+\frac{a(a+b)}{a+b-1}$, an inequality that reduces to the true inequality $b<1$. Hence, indeed, $\pi^*(1)=1$, and so Kantians successfully resist invasion by Nashers.

Summarizing, the only regions in which Kantians successfully resist Nash invasion are Va and VII. But these regions are identical, for note that

$$\text{Region Va} = N_2 \cap K_2 \cap \{a+b<1\} \cap \{b<1\} =$$
$$\{a<0\} \cap \{a+b<1\} \cap \{b<1\} = \{a<0\} \cap \{b<1\} = N_4 = \text{Region VII} \quad .$$

The reason these appeared as separate regions in the analysis is because there are multiple Nash equilibria in this region -- in Region Va we assumed Nashers play the equilibrium strategy $q^*=0$ and in Region VII they play the equilibrium strategy $q^*=q_1^*$. ■

We now have:

Corollary 8.5

For the class of (a,b) *games, the following conditions are equivalent:*

a. $a<0$ and $b<1$.

b. Kantian players drive Nash players to extinction.

c. The game is one of pure coordination.

d. The game is supermodular and a mixed-strategy Nash equilibrium exists.

Proof:

$a \Leftrightarrow b$ by Proposition 8.4. $a \Leftrightarrow c$ by lemma 8.2. $a \Leftrightarrow d$ by lemma 8.3. ■

In words, Kantian optimizers possess an evolutionary advantage over Nashers precisely when the game is one of pure coordination. (If the Nashers were able to coordinate on the good equilibrium, then they and Kantians would appear, in these games, to be identical.) These games are all ones with strategic complementarity – the property that cooperation begets cooperation – but strategic complementarity alone is insufficient to guarantee that Kantians drive Nashers to extinction (part *d*). In addition to supermodularity, the existence of a mixed-strategy Nash equilibrium implies (by lemma 8.1) that no pure-strategy symmetric Nash equilibrium exists. If (X,X) were a pure-strategy Nash equilibrium then Nashers and Kantians would coexist and be indistinguishable, and if (Y,Y) were the unique symmetric Nash equilibrium, then Nashers drive Kantians to extinction. This is why the existence of a mixed-strategy NE must be appended to supermodularity in part *d*.

Note the Prisoners' Dilemma (PD) games comprise the region $\{a<0\}\cap\{b>1\}$: Nashers drive Kantians to extinction in PD games, whether it is a game where the SKE is X or p_1^*.

8.3 Summary

In games of pure coordination, Kantians drive Nashers to extinction. These are games with strategic complementarity, so cooperation begets cooperation. In such games, Kantians can survive without punishing those *ex post* who play strategy Y. Indeed, the advantage to cooperating is sufficiently strong that Nashers do not survive, even though they are not punished for playing autarchically.

But if invasion by Nash players is a threat, and our result shows it will be if the game is not a coordination game, then Kantians must either learn to recognize Nashers, or to punish them *ex post*. Punishment *ex post* is properly targeted when one knows that an opponent who plays Y must be a Nasher, and this is the case when the simple Kantian

equilibrium is the pure strategy *X*. This is the case in a class of PD games: see proposition 2.2. If such games are prevalent, then we can expect that, if Kantians survive, they will have evolved to punish Nashers. For otherwise, if that kind of PD game is typical of real life, we would not observe Kantian behavior. In laboratory games, it is usually the case that it is impossible to identify the type of one's opponent, and one sees, ubiquitously, that non-cooperators are punished by cooperators, in games where the Kantian equilibrium is to cooperate with probability one. Indeed, Boyd et alii (2003) provide an argument based on group selection for how 'altruistic' punishment can evolve[4]. They conclude that 'group selection can maintain altruistic punishment and altruistic cooperation over a wider range of parameter values than group selection will sustain altruistic cooperation alone (p. 3533).'

Bowles and Gintis (2004) give the following example. Farmers in the Indian village of Palanpur can plant their seeds early or late. The farmers are caught in a bad Nash equilibrium, planting late. It's a bad equilibrium because the harvests are relatively small with late planting. However, if a (sole) farmer deviates from the late-planting equilibrium and plants early, the birds will eat his seed. It is also a symmetric Nash equilibrium to plant early: if all do this, the birds will take relatively few seeds from each plot, and the farmers enjoy a more abundant harvest than if all planted late. The early-planting equilibrium Pareto dominates the late-planting equilibrium. This is a pure coordination game, and the simple Kantian equilibrium is that all plant early. But lacking the Kantian optimization ethos, Palanpur is stuck in the bad Nash equilibrium.

In contrast, Boehm (2012) describes a band of !Kung bushmen who hunt for game by spreading their large nets at the edge of the forest. Women and children beat the bushes in the forest, scaring the game, which run into the nets. One day, Cephu secretly places his net in front of the nets of the other hunters, and catches a large fraction of the game. Over the next few days, he is criticized by others, and ostracized. He is threatened with expulsion from the band. Soon, he takes the game he caught and shares it with the others. He cries and swears not to do this again. The question we must ask is

[4] Punishment, in environments such as the one in this chapter, is biologically altruistic because the individual receives no return for punishing his opponent, an act that is presumed to be costly to him.

whether Cephu is one of a small number of Nashers in a community that consists mainly of Kantians, or whether everyone is a Nasher, kept in line by the threat of expulsion from the community as a punishment for non-cooperative behavior. According to Boehm's account, the punishment does not occur immediately. It takes time for one hunter to raise his criticism of Cephu. It does not appear as if the first one to speak out against Cephu had a responsibility to do so, and would have been punished by others had he failed to speak out. It does not appear from the account that his speaking out was part of a Nash equilibrium in a multi-stage game. But when he speaks out, others join the attack on Cephu .

Distinguishing the motives for cooperation between these two possibilities (that everyone is a Nasher, but that most are deterred from selfish behavior by the threat of punishment, versus the situation in which only a few Nashers exist but most are Kantians) is important but difficult. It seems to me natural to conjecture that the first hunter who criticized Cephu was a Kantian. He initiated costly punishment when others were holding back. Other hunters were conditional Kantians, who undertook punishing action once the ball got rolling. Boehm describes how mild punishments usually succeed in controlling non-cooperators. But sometimes they do not, and the non-cooperators are either expelled from the band or, if their behavior is really costly to the group, executed. These responses to non-cooperators in egalitarian bands of hunter gatherers appear to be ubiquitous around the world.

Words: 4443

Chapter 9. Alternative approaches to cooperation

9.1 The traditional approach

The traditional approach to explain cooperation is to model it as a Nash equilibrium in a repeated game where players are self-interested, but will be punished by others if they fail to play the 'cooperative' strategy. Since punishment is postulated to be costly to the punisher, it must also be the case that those who fail to punish non-cooperators are themselves punished by others in the next round of the game. Cooperation by all players comprises a Nash equilibrium only if the game has an infinite or unknown number of stages. For if there were a known last stage T, then at stage $T-1$, players would fail to cooperate, because no rational self-interested player would punish those non-cooperators at stage T. Hence cooperation unravels.

As I said earlier, I find this explanation of cooperation unconvincing as a general explanation of the many examples of cooperation that we observe. My objection is intuitive: I cannot believe that examples of human cooperation are maintained solely through the fear of punishment. It is much more reasonable, to me, to think that many people have internalized the norm of cooperation, and they would cooperate, absent punishments. However, punishments are needed to keep the autarkic optimizers (the Nash players) in line. Indeed, the literature is full of examples of repeated games with a known number of stages (T) in which cooperation occurs at the early stages (at least), although the Nash equilibrium of the repeated game with self-regarding preferences is a complete failure of cooperation. Moreover, when punishment is possible, a large fraction of players cooperate for the whole game – and those who fail to cooperate at stage $T-1$ are often punished by others at stage T, even though this is inconsistent with autarkic optimization (see for example Fehr and Gintis (2007)).

Gintis (2000) denotes the cooperation that is enforced by punishment 'weak reciprocity,' and he contrasts this with his own theory of 'strong reciprocity.'

9.2 Strong reciprocity

Gintis (2000) proposes that 'a strong reciprocator is predisposed to cooperate with others and punish non-cooperators, even when this behavior cannot be justified in terms

of self-interest, extended kinship, or reciprocal altruism.' Later , in Bowles and Gintis (2011), the authors write: 'We commonly observe that people sacrifice their own payoffs in order to cooperate with others, to reward the cooperation of others, and to punish free-riding, even when they cannot expect to gain from acting in this way. We call the preferences motivating this behavior *strong reciprocity* (p. 20).'

Gintis often speaks of the propensity to take actions that are not Nash-best responses in a game as *altruistic*. I believe there is a distinction that should be maintained between the concepts of *biological* and *psychological* altruism. Gintis is using the term in its biological sense: any action that an individual member of a species takes that helps others at a cost to itself is considered altruistic in biology, regardless of the cause of the action. In contrast, psychological altruism is what economists typically mean by the term – that an individual's preferences assign positive value to the welfare of others, and hence these costly actions that help others are motivated by a *desire* to help. Because Gintis uses the term in its biological sense, he denotes any action that objectively helps others at a cost to oneself as being the consequence of maximizing non-self-regarding preferences. But this does not follow. I may punish a non-cooperator because I am offended that he has broken a norm – to behave cooperatively -- not because I care about the welfare of others. Equating such motivations with having altruistic preferences in the psychological sense is, I believe, a confusion.

In contrast, the preferences of players in all chapters of this book except chapter 5 are self-regarding, yet nevertheless the Kantian equilibrium typically involves cooperation. I achieve this by asserting that players use an optimization protocol that is not Nash's, and doing so is *not* motivated by psychological altruism, but by understanding that, in a situation of solidarity, players must 'hang together or hang separately.' Because Gintis has no option of varying the optimization protocol, he is forced to explain cooperation by saying that players have non-self-regarding preferences. There is a social norm in my story, but it is not modeled as an argument of preferences, but rather induces the choice of optimization protocol.

Is this a distinction without a difference? I believe not. The evidence for my claim is that the theory of strong reciprocity does not tell us how cooperation is achieved except in very simple cases, where the cooperative strategy profile is *obvious.* In contrast,

the theory of Kantian optimization, is *explicit* about the optimization program of each player, and hence can define a concept of equilibrium even in 'complex' games such as the production economies that we have used to study common-pool resource problems. In particular, players may possess different preferences. The cooperative equilibria, for example, in production economies that I called 'fishing' economies, where the allocation rule is proportional, are far from being ex-ante obvious. They are, however, multiplicative Kantian equilibria with standard, self-regarding preferences. In contrast, in chapter 6, I showed that to achieve these Pareto-efficient allocations as Nash equilibria of a game where players have extended preferences over the entire allocation is mathematically possible, but unconvincing, as it would require a theory of how players adopt the 'right' non-self-regarding preferences. For those extended preferences are *in general* not given by any simple social welfare function whose arguments are the players' utilities.

Fehr and Gintis (2007) advocate strong reciprocity to explain the results of experiments with a public-good (PG) game. The are N players. Each is endowed with an amount of resource Y. Each can contribute any amount $0 \le y_i \le Y$ to a common pot. The pot is multiplied by a number M by the experimenter, where $1 < M < N$. This expanded pot is then divided equally among all players. Thus the payoff to a player i is:

$$Y - y_i + \frac{M}{N}\sum_j y_j \ . \tag{9.1}$$

The Nash equilibrium (where the strategies are $\{y_i\}$) in the one-shot game is that all players contribute zero. Since the game is symmetric, it is appropriate to look at the simple Kantian equilibrium. Each player, under the simple Kantian protocol, maximizes $Y - y + \frac{M}{N}Ny = Y + (M-1)y$ under the constraint $0 \le y \le Y$, and the solution is $y = Y$ because $M > 1$.

The one-shot game is, however, not the one that Fehr and Gintis (2007) describe, but rather a repeated game with two treatments – either with, or without, the option to punish those who fail to cooperate. The experimental result is that, without the possibility of punishment, cooperation is quite high in the beginning stages, but deteriorates by the last stage (10) to very little. With the possibility of punishing non-

cooperators, cooperation becomes virtually complete by the last stage, and some players in the last stage punish those who failed to cooperate in the penultimate stage.

How do I explain the tendency for 'altruistic punishment' in one-shot games (like the ultimatum game) or in the last stage of a repeated game? In chapter 8, I observed that Kantian players would be driven to extinction if they did not learn to punish non-cooperators who are present in their population. Thus, the tendency to punish must have evolved if we are to observe cooperation in mixed populations. (As I wrote, Boyd et al (2003) propose an evolutionary mechanism.)

Suppose one wishes to rationalize cooperation in the one-shot PG game of Fehr and Gintis (2007) by *psychological* altruism. Suppose that players append to the payoff function in (9.1) a utilitarian social welfare function, so that player *i* maximizes:

$$V^i(y_i;y_{-i}) = Y - y_i + \frac{M}{N}\sum_{j=1}^{N} y_j + \alpha^i \sum_{k\neq i}\left(Y - y_k + \frac{M}{N}\sum_{j=1}^{N} y_j \right) . \qquad (9.2)$$

We have $\frac{\partial V^i}{\partial y_i} = -1 + (1+\alpha^i)\frac{M}{N}$. It therefore follows that the Nash equilibrium of this game is:

$$y_i = \begin{cases} Y, \text{ if } (1+\alpha^i)\frac{M}{N} > 1 \text{ or } \alpha^i > \frac{N-M}{N} \\ [0,Y] \text{ if } (1+\alpha^i)\frac{M}{N} = 1 \\ 0, \text{ if } (1+\alpha^i)\frac{M}{N} < 1 \end{cases} .$$

The values for the Fehr-Gintis experiment are $M = 2, N = 10$. Thus altruists would have to have $\alpha^i > 0.8$ in order to be play fully 'cooperatively.' This strikes me as unreasonably high as a credible explanation of cooperation.

9.3 Conditional cooperation

In reality, I think there are very few individuals who always use the Nash protocol or who always use the Kantian protocol. Most people are *conditional Kantians.* I propose that each person *i* has a threshold, q_i , and will optimize using the Kantian

protocol if and only if she observes that fraction q_i of the relevant population (set of players) is cooperating. Figure 9.1 plots a typical cumulative distribution function of thresholds in a population.

[figure 9.1 here] [figures for this chapter to be appended later]

In the case of figure 9.1, the stable behavior is that fraction q^* of the population will play the cooperative strategy. For suppose fraction $q < q^*$ were cooperating; then the fraction who desire to cooperate, given cooperation at level *q*, is greater than *q*, so *q* will increase. A similar argument shows that no value $q > q^*$ is stable.

[figure 9.2 here]

In Figure 9.2, there are three equilibrium cooperation frequencies, q_1^*, q_2^* and 1 . Equilibrium q_1^* is stable – a slight displacement from it will induce a dynamic returning to q_1^*. But q_2^* is unstable. If a shock causes the number of cooperators to fall, the new equilibrium will be at q_1^* . If, perchance, a shock increases the frequency of cooperation, the new equilibrium will be at $q^* = 1$.

[fig 9.3 here]

In many situations of solidarity, trust must be established to build cooperation. Often, there is a small core of individuals who are *saints* – their threshold is $q = 0$. In Figure 9.3, I have drawn a distribution function of thresholds where fraction *f* of the population are saints. Saints induce others, whose thresholds are strictly positive, to cooperate. Cooperation builds as those who are increasingly skeptical join – their trust of others increases as they see others joining the movement. Eventually, at q^*, the limit is reached. For the game of recycling, q^* is now quite high in many cities, even though there is scarcely any punishment or ostracism of those who fail to recycle. For the voting game, q^* varies across countries.

Of course, the dynamic described here can describe conditional strong reciprocators as well as conditional Kantians.

9.4 Rabin and the kindness function

Matthew Rabin has proposed an explanation of cooperation, which, I think, could provide micro-foundations for Gintis's strong reciprocity. A relatively simple description of the approach is found in Rabin (2004). Strictly speaking, the approach is not a special case of Nash equilibrium, because players' actions depend not only on the actions of others, but on their beliefs about the motives of other players. Roughly speaking, players wish to be generous to other players who, they think, are motivated by being kind to them, and they wish to punish players who , they think, are motivated by being unkind. This produces extended utility functions, where arguments incorporating these behaviors are included . A *fairness equilibrium* is then defined as what Rabin (2004) describes as ' the analog of Nash equilibrium for psychological games.'

I have three reservations about fairness equilibrium. My first reservation is that, when all is said and done, fairness equilibrium involves autarkic reasoning, as Rabin says when he writes that it is the analog of Nash equilibrium in psychological games. I return to my deepest motivation for the Kantian approach, which is that I think a fundamentally different kind of optimization protocol is needed to explain cooperation. Lest this reservation be viewed as too demanding or radical, let me remark that Kantian optimization *is* within the genre of what Jon Elster calls methodological individualism. That is, it provides an explanation of cooperation at the level of individual choice. A *truly* radical critique would argue that cooperation is an 'emergent' phenomenon that cannot be explained at the individual level, a view to which I surely do not subscribe.

My second reservation is that fairness equilibrium assumes considerable sophistication on the part of players. It is required that higher order beliefs match actual behavior. Thus, a player's payoff depends not only on the strategy profile, but on his beliefs about the other player's strategy choice, and his beliefs about the other player's beliefs about his strategy choice. Now the complexity of reasoning required could be invoked to explain why cooperation is difficult to establish in reality; but I would rather say that cooperation is prevalent in reality, and a simpler explanation would be more convincing.

My third reservation is that it is unclear how the theory could be extended to more complex economic environments than 2×2 games. As with strong reciprocity, it appears to require an ex-ante conception of what constitutes cooperative (or kind or fair)

play. As I have been at pains to establish, it is often, in real life, not clear what the cooperative action is. In the fishing economy, how can I decide whether another fisher's labor supply is being kind or unkind to me?

9.5 *Homo moralis* (Alger and Weibull, 2013)

Alger and Weibull propose a model in which agents are randomly matched in pairs, from an infinite population, and play one-shot games, where strategies are chosen from an abstract (compact, convex) strategy space X. The 'material payoff' to a player is a function $\pi(x,y)$ of the strategy profile, where the function π varies across players. However, the *utility* function of the player is:

$$u_\kappa(x,y) = (1-\kappa)\pi(x,y) + \kappa\pi(x,x), \tag{9.3}$$

some $\kappa \in [0,1]$. Here is the authors' interpretation of (9.3):

> It is as if *homo moralis* is torn between selfishness and morality. On the one hand, she would like to maximize her own payoff. On the other hand, she would like to "do the right thing," that is, choose a strategy that, if used by all individuals, would lead to the highest possible payoff. This second goal can be viewed as an application of Kant's (1785) categorical imperative, to "act only on the maxim that you would at the same time will to be a universal law." Torn between these two goals, *homo moralis* chooses a strategy that maximizes a convex combination of them. If $\kappa = 0$, the definition of *homo moralis* coincides with that of "pure selfishness," or *homo oeconomicus*; given any strategy opposite extreme, $\kappa = 1$, the definition of *homo moralis* coincides with that of "pure morality," or *homo kantiensis*; irrespective of what strategy the other party uses (or is expected to use), this extreme variety of *homo moralis* will use a strategy in $\arg\max_{x \in X} \pi(x,x)$. (p. 2276)

The focus of their article is the study of when agents with preferences like *homo moralis*, can survive invasion by other agents. Under stipulated conditions, including assortative matching of players in the 2×2 games, they argue that such agents can successfully

resist invasion. The equilibrium in the game played by two players (with preferences of this sort) is a Nash equilibrium (actually, Bayesian Nash). So the Kantian aspect of the model is embedded in the utility function of (9.3), not in the optimization protocol.

To compare this approach to mine, let's begin with the case when the two players are identical, and when they are both type *homo kantiensis* – that is, $\kappa = 1$. Then it is trivial to observe that (x^*, x^*) is a Nash equilibrium of the game where each maximizes $\pi(x,x)$, where x^* is the (unique) maximum of $\pi(x,x)$. (We have observed this earlier in Proposition 6.1.) But the similarities between the two approaches disappear when we look at players of different types.

When player types are different, as we have seen, simple Kantian equilibrium, which corresponds to *homo kantiensis*, is the wrong way to model cooperation. We must replace it with multiplicative or additive or some other 'Kantian variation.' Let's write down the definition of Nash equilibrium corresponding to (9.3), and $K^{\times}$ equilibrium, in terms of best-response functions or correspondences. For traditional Nash equilibrium, using the notation of Alger and Weibull, the best-response functions are:

$$\begin{aligned} \beta^1(y) &= \arg\max_x \pi^1(x,y) \\ \beta^2(x) &= \arg\max_y \pi^2(x,y) \end{aligned} . \tag{9.4}$$

For Alger-Weibull equilibrium (minus the Bayesian part), the best-response functions are:

$$\begin{aligned} \hat{\beta}^1(y) &= \arg\max_x (1-\kappa)\pi^1(x,y) + \kappa\pi^1(x,x) \\ \hat{\beta}^2(x) &= \arg\max_x (1-\kappa)\pi^2(x,y) + \kappa\pi^2(x,x) \end{aligned} . \tag{9.5}$$

For $K^{\times}$ equilibrium, the best-response functions are:

$$\begin{aligned} \tilde{\beta}^1(x,y) &= r^1 x, \text{ where } r^1 = \arg\max_r \pi^1(rx,ry) \\ \tilde{\beta}^2(x,y) &= r^2 x, \text{ where } r^2 = \arg\max_r \pi^2(rx,ry) \end{aligned} \tag{9.6}$$

.

For the three cases, an *equilibrium* is a strategy pair (x^*, y^*) such that :

$$(x^*, y^*) \in (\beta^1(y^*), \beta^2(x^*)) \text{ or } (x^*, y^*) \in (\hat{\beta}^1(y^*), \hat{\beta}^2(x^*)) \text{ or } (x^*, y^*) \in (\tilde{\beta}^1(x^*, y^*), \tilde{\beta}^2(x^*, y^*)), \tag{9.7}$$

respectively. The key difference between the first two cases and the third is that the first two cases use a best-reply function (or correspondence) that is a mapping $X \to X$, while Kantian equilibrium uses a mapping defined $X^2 \to X$. (If there were *n* players, the best-reply correspondences for Nash equilibrium are mappings $X^{n-1} \to X$, and the best-reply correspondence for Kantian equilibrium is a mapping $X^n \to X$.) This apparently small formal difference, however, is key: for it expresses the idea that a Kantian player thinks of a counterfactual where *all players* vary their strategies in the same way.

In my view, the Alger-Weibull approach is still wedded to the idea that cooperation can be thought of as a non-cooperative (Nash) equilibrium, and this is fundamentally different from the view I advocate, that we must conceptualize cooperation as involving a fundamentally different kind of optimization. For a graphical presentation of the difference between (9.4) and (9.6) , recall Figure 4.1. Kantian optimizers imagine a counterfactual that lies in the same ray, while Nash optimizers think of counterfactuals lying on different rays.

9.6 Surveys and interviews

In 2014, several questions about recycling were included on the Innovation Sample of the German Socio-Economic Panel (SOEP-IS). I summarize the responses to these questions[1]. 1495 people responded to these questions: 96% (1435) said that they recycle plastic and paper. Those who recycle gave the reasons for recycling that are tabulated in Table 9.1. The choices were fixed (close ended question), and a respondent could choose more than one reason.

[1] The responses were sent to me by Carsten Schröder of SOEP in DIW, Berlin. I am grateful to Carsten for including these questions on the survey.

Reason	**Number respondents**
1. Every little bit counts	811
2. It's what a person should do	1016
3. I would feel guilty otherwise	461
4. I'd like everyone to recycle	466
5. It keeps the house neater	464
6. People in my neighborhood do	175
7. Waste collection charges are cheaper	374

Table 9.1 German Social Survey: reasons for recycling, close-ended

The clearest Kantian reasons are the second and the fourth. The first reason could be thought of as a Nash response, where even the marginal contribution of the individual to a green environment is *greater* than the personal cost of recycling. The fifth reason is Nash. The seventh response could either be interpreted as Kantian, if reducing waste-collection charges is a public good to which everyone should contribute by recycling; or it is a confused answer, if the respondent believes her own recycling will lower her charges. Certainly the responses are consistent with a large fraction of Kantian optimizers, although there can be many interpretations.

The 4% of respondents who do not recycle provided the reasons listed in Table 9.2 (again, more than one reason could be given).

Reason	**Number**
1. Would not make a meaningful difference	21
2. No reason to recycle	13
3. I don't care	11

4. Cost of recycling not worth it	2
5. I don't have the time	12
6. People in my neighborhood don't	17
7. Waste collection charges aren't much higher if I don't	9

Table 9.2 German Social Survey: reasons for not recycling, close-ended

Most of these reasons can be explained by Nash optimization.

To the question "How many people do you think recycle in your neighborhood?" 75% of respondents answered two-thirds or more, with 48% responding 95%. Among those who do not currently recycle, one-half said they would do so if a higher fraction of people recycled. This is consistent with these 28 people being conditional Kantians with high thresholds (or having perceptions of a low participation rate).

In 2014, I conducted a small survey using Amazon's Mechanical Turk. To the open-ended question "Why do you recycle?", I classify the responses as reported in Table 9.3. (Several respondents gave more than one reason.)

Reason for recycling	**Number**
1. Helping the environment, save the earth, etc.	25
2. It's easy to do	13
3. Dislike of waste	11
4. Save money, lower my trash bill	7
5. Be a part of making a difference	7
6. It's the right thing to do, duty	5
7. Feel guilty if I didn't	2
8. For future generations	1

Table 9.3 Mechanical Turk: Open-ended reasons for recycling

The first reason is certainly not Nash; it could be Kantian. It is similar to the fifth reason, although 'being a part of making a difference' is more clearly Kantian. The fifth reason is also called indicative of an expressive utility from participation. Reason 8 is clearly altruistic, although reason 1 may be so as well.

To the question "Why do you not recycle?" almost all answers were of the form that recycling is too costly. One person, with the clearest Nash consciousness, wrote 'because my action would not make a meaningful difference.'

Close-ended responses to the question "Why do you pay your taxes fairly honestly?" are tabulated in Table 9.4.

Reason for paying taxes honestly	**Number**
1. Fear of being caught, paying a fine	52%
2. Country needs the revenue	20%
3. It's what everybody should do	58%
4. Not worth the worry if I didn't	44%
5. What if nobody paid their taxes?	12%

Table 9.4 Mechanical Turk: Close-ended responses to paying taxes

Reason 1 is Nash. (There is a large literature arguing that the probability of being fined and the sizes of fines are far too small to generate the degree of tax compliance that is observed in the United States. See, for example, Frey and Torgler [2007].) Reasons 2,3 and 5 are consistent with Kantian reasoning.

In some treatments, respondents were given an open-ended question about tax compliance. The responses are given in table 9.5.

Reasons for tax compliance	**Number**
1. It's the right thing to do; duty as citizen; don't want to cheat the gov't; part of social contract; everyone does their fare share	16

2. Fear of audits, not worth risk	28
3. Support my gov't; pay for essential services; tax money goes to good use	7
4. Because it's the law	5

Table 9.5 Mechanical Turk: Open-ended question on tax compliance

Reasons 1 and 3 seem Kantian; reason 2 is Nash.

Answers to the close-end question "Why do you vote in national elections?" are given in table 9.6.

Reason for voting	**Number**
1. Duty of a citizen	68%
2. Like to participate in choosing	57%
3. My vote could make a difference if close	46%
4. It doesn't take much time	19%
5. If I didn't vote, why should anyone?	8%

Table 9.6 Mechanical Turk: close-ended reasons to vote

Reason 1 is a social norm, which could be interpreted as Kantian. Reason 2 is the expressive reason for voting, in which participation has a value. Reason 3 is Nash. Reason 5 is Kantian.

To the open-ended version of this question, people responded with reasons tabulated in table 9.7.

Reasons to vote	**Number**
1. Every vote makes a difference; want my voice to count; have a say	20

2. It's my duty; people died for the right to vote	4
3. Voting gives me a right to complain	3
4. Although one vote may not make a difference, if everyone believed this…	1

Table 9.7 Mechanical Turk: Open-ended reasons to vote

The vast plurality in table 9.7 seem to view voting as expressive (reason 1). Only reason 4 is clearly Kantian. Both the close-ended and open-ended questions in the Mechanical Turk survey indicate a small fraction of Kantian optimizers in regard to voting.

Stefan Penczynski of the University of Mannheim has conducted laboratory experiments with public-good games where players consist not of single agents but of teams of two. Before the play of the game, the members of each team discuss with each other (but not with other teams) how much their team should contribute to the public good. Then one member from each team is randomly chosen to play the game. The purpose of this set-up is to elicit from individuals the reasons behind their strategy choice. Here is a summary of the reasons that players gave in their pre-play communication with their team partners[2].

1. "If everybody contributed the maximum, that would be the most rational and best decision for all. However, I doubt that everybody will do so; that is why, in this case, I would rather keep everything and not be stupid at the end."

2. "I just hope that most people follow the principle of maximizing total welfare. Obviously, we could just all our points and then hopefully get all the Taler from the other teams. But I am human in my decision."

3. "It makes sense here to think strategically because one's own Taler are not as valuable as the ones the other teams send over. Hopefully the other teams think strategically as well."

[2] Personal communication with S. Penczynski.

4. “Hold everything or pass everything? It would be better to pass everything , if the other teams think this way as well.”
5. “I would pass everything as this way the total payout is increased. I assume that the others will pass everything as well.”
6. “ I think we should rely on the others. Holding Taler leads to much too small an amount. That is the trick: If everybody goes along and passes the Taler, we can leave with more money. And I think we should be optimistic here.”
7. “I trust that the other team sends us everything as well – this way everybody gets the maximal amount! We have to trust though; without this a market does not work.”
8. “So basically the money amount for everybody will be maximized when everybody passes all Taler. When everybody thinks like that, we would earn 5 Euro. But there will be those that hope to get everything donated without doing something for it themselves. That would not be good, but I hope that we meet nice people that give us everything as well.”
9. “I am afraid that few will play cooperatively in order to raise the total profit of all participants. It would be ok for me, however, to pass 40 Taler in order to make the most of the available money for all.”

What appears ubiquitous in these discussions is the willingness to play the Kantian strategy. The doubts raised are due to lack of trust. None of the players expressed the Nash argument. Several players rationalize the cooperative strategy by referring to the maximization of total profit. As I pointed out in chapter 6, this ambiguity always exists in symmetric games with identical payoff functions : that is, the Kantian equilibrium of the game with self-interested preferences is also a Nash equilibrium of the game that maximizes total the total payoff.

9.7 The Acequias of northern New Mexico

(to be written, perhaps)

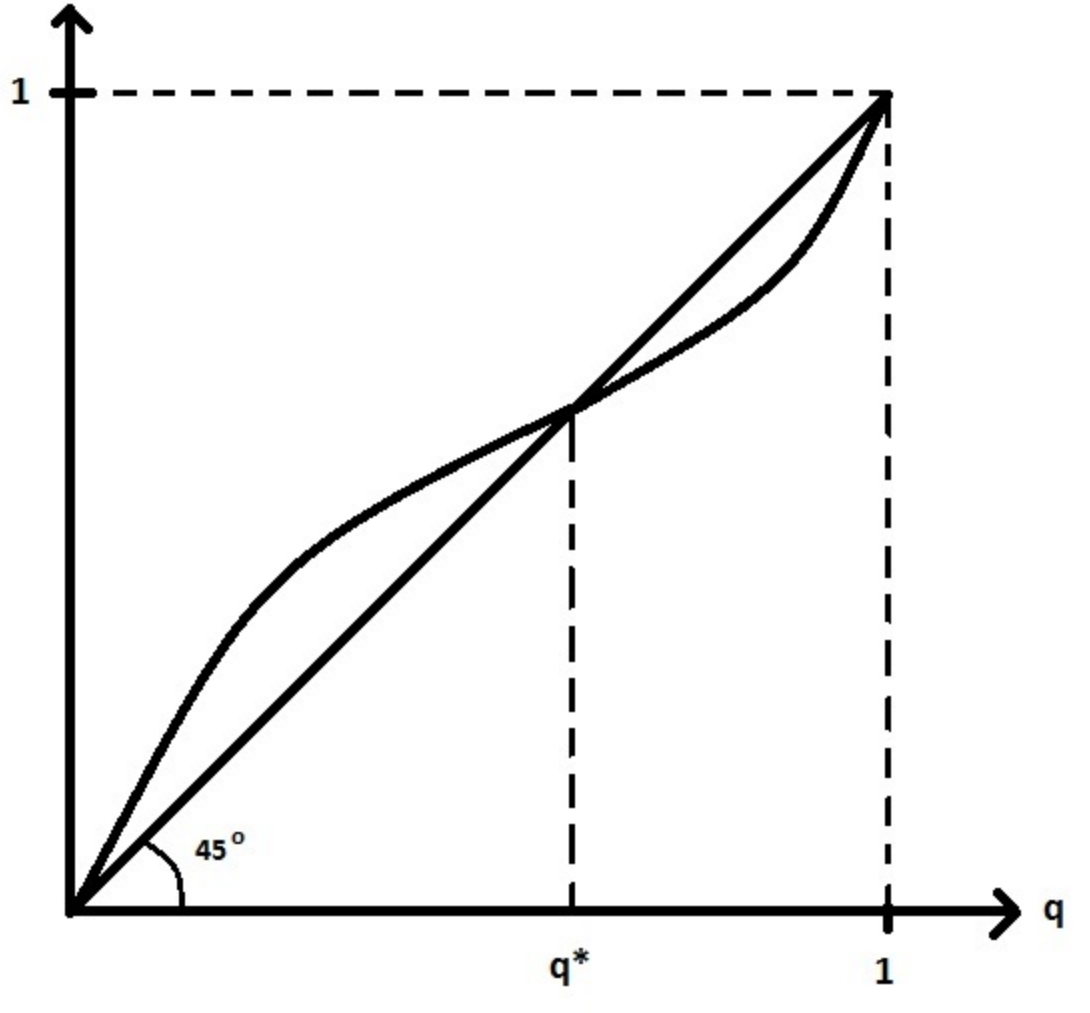

Figure 9.1

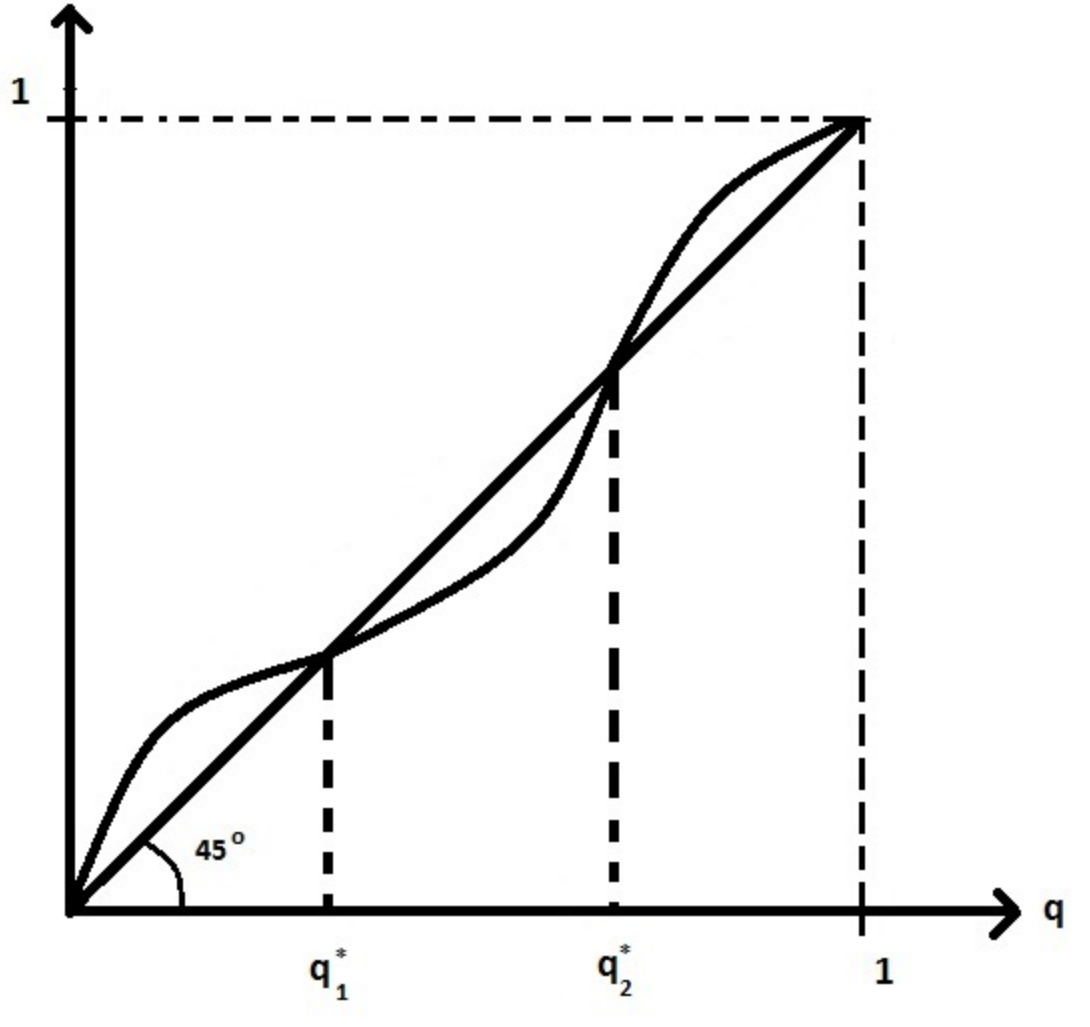

Figure 9.2

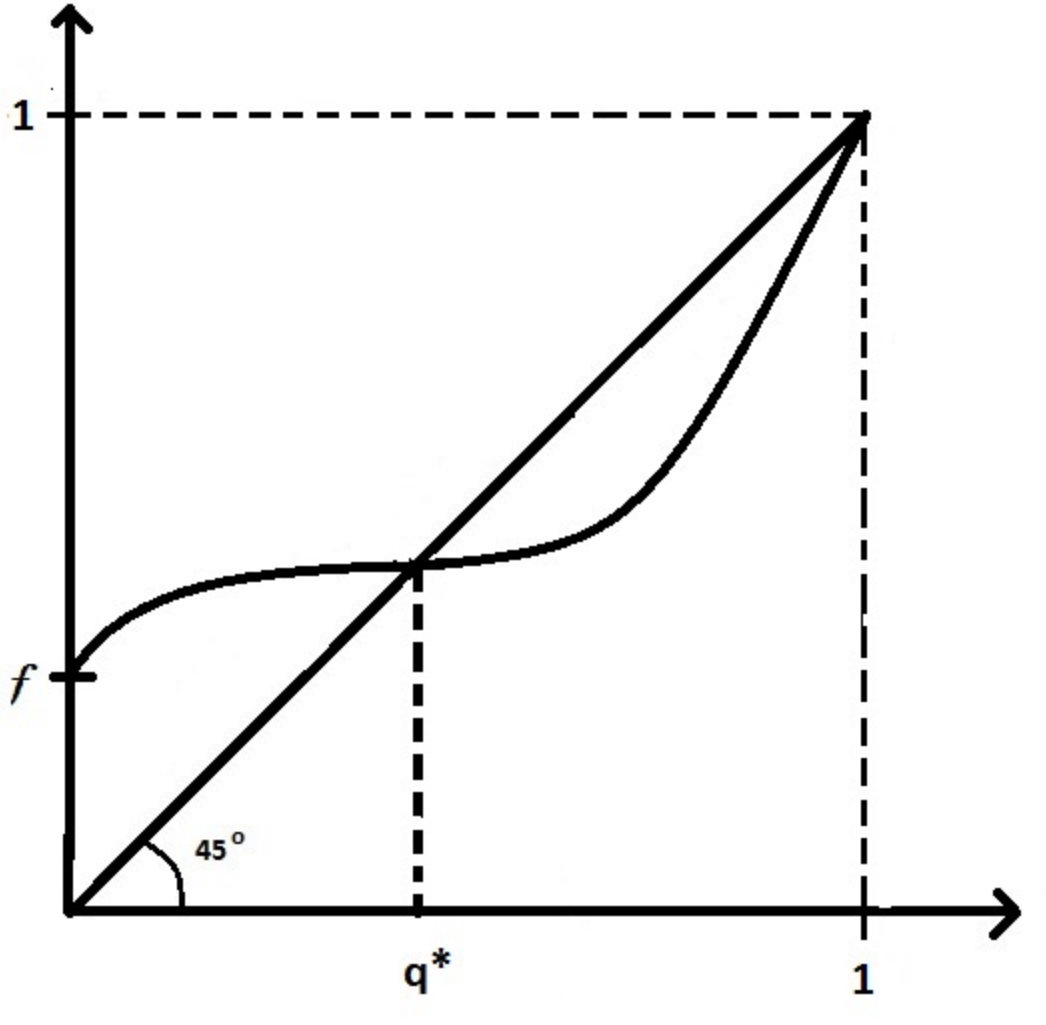

Figure 9.3

Chapter 10 A generalization to more complex production economies

With a few exceptions, the strategies that players employ in the games in this book are unidimensional. In particular, in the production economies studied, production is a function of a single kind of efficiency labor or effort. In this chapter, I show that there are limited generalizations of Kantian optimization to the case where production is a function of several kinds of labor. Thus, I now assume that the production function G maps $\Re_+^l$ into $\Re$: there are l types of labor/effort. An aggregate vector of labor supplies will be denoted $\mathbf{E} = (E^1,...,E^l)$; each agent will supply only one kind of labor, but there may be many agents supplying each kind of labor. We denote by $l(i)$ the type of labor that agent i supplies: thus, $l(i) \in \{1,2,...,l\}$ for every i. Although the labor vector is multi-dimensional, each player's strategy continues to be unidimensional.

Conceptually, it seems clear why we must maintain the restriction of unidimensionality for the individuals' strategies (effort supplies). The idea behind Kantian optimization is that there is a natural conception of what it means to take the same kind of action. If actions were multi-dimensional, it is difficult to conceptualize what the same kind of action would be. Here, we show that, as long as individuals each have unidimensional effort strategies, even if those strategies are drawn from different sets (in the sense of involving different *types* of labor), players can cooperate by using the familiar multiplicative and additive conceptions of Kantian variation. There will, however, be restrictions on the production functions G required to derive efficiency results.

We now suppose that there are n workers, each with a concave, differentiable utility function $u^i(x,E)$, where x is consumption and E is the amount of the unique type of effort/labor that agent i is capable of supplying. The types of labor that agents can supply are specified by a (single-valued) function $l:\{1,2,...,n\} \to \{1,2,...,l\}$. Thus, a feasible allocation of effort is a vector $(E_1^{l(1)},E_2^{l(2)},...,E_n^{l(n)})$ in $\Re_+^n$. The associated vector of aggregate labor supplies is $\mathbf{E} = (E^1,...,E^l)$ where $E^j = \sum_{i \text{ s.t. } j=l(i)} E_i^{l(i)}$ for $j = 1,2,...,l$. The vector of consumptions $x = (x_1,...,x_n)$ must satisfy $\sum_i x_i \leq G(\mathbf{E})$.

Definition 10.1 A production function $G:\Re_+^l \to \Re_+$ is *homothetic* if it is differentiable and for all pairs of components j,k there is a constant γ_{jk} such that , for all positive α and for all vectors $\mathbf{E} \in \Re_+^l$, $G_j(\alpha\mathbf{E}) = \gamma_{jk} G_k(\alpha\mathbf{E})$, where G_j denotes the j^{th} partial derivative of *G*.

Indeed, we have $\gamma_{jk} = \dfrac{G_j(\mathbf{E})}{G_k(\mathbf{E})}$. If $l = 2$, this is just the familiar condition that along any ray (emanating from the origin), the slopes of the isoquants of *G* are constant. More generally, the tangent hyperplanes to the isoquants of *G* are parallel along any ray in $l-$ space.

Definition 10.2[1] A production function $G:\Re_+^2 \to \Re_+$ is *C-homothetic along a path* $(\theta_1,\theta_2) \in \Re_{++}^2$ if it is differentiable, and for any vector $\mathbf{E} = (E^1,E^2) \in \Re_+^2$, the slopes of isoquants of *G* are constant along expansion paths $(E^1 + \alpha\theta_1, E^2 + \alpha\theta_2), \alpha \geq 0$. In other words, $G_1(E^1 + \alpha\theta_1, E^2 + \alpha\theta_2) = \gamma G_2(E^1 + \alpha\theta_1, E^2 + \alpha\theta_2)$ for any non-negative α , and so the constant γ has the value $\dfrac{G_1(\mathbf{E})}{G_2(\mathbf{E})}$.

I denote this property by '*C*-homotheticity' because it was first used by Chipman (1965), but for utility functions[2]. Consider the concave production function $G(E^1,E^2) = 2 - e^{-\theta_2 E^1} - e^{-\theta_1 E_2}$. Note that the slope of the isoquant of this function at (E^1,E^2) is $-\dfrac{\theta_1 e^{-\theta_1 E^2}}{\theta_2 e^{-\theta_2 E^1}}$, which is constant if $\theta_2 E^1 - \theta_1 E^2 = k$, or along expansion paths of slope $\dfrac{\theta_2}{\theta_1}$ in the (E^1,E^2) plane.

---

[1] I state the definition for $l = 2$, to avoid discussion of tangent planes to iso-level surfaces of *G*, which would be needed for higher dimensions.

[2] The citation to Chipman (1965) is thanks to J. Silvestre.

Proposition 10.1

A. *Let* $l = 2$, *let G be twice differentiable, and let* $\mathbf{H}$ *be the Hessian matrix of G. Then G is homothetic if and only if, for all* $\mathbf{E} = (E^1, E^2)$:

$$(E^1, E^2)\mathbf{H}\begin{pmatrix} G_2 \\ -G_1 \end{pmatrix} = 0 \ . \tag{10.1}$$

B. *G is C-homothetic along a path* (θ_1, θ_2) *if and only if, for all* $\mathbf{E} = (E^1, E^2)$:

$$(\theta_1, \theta_2)\mathbf{H}\begin{pmatrix} -G_2 \\ G_1 \end{pmatrix} = 0 \ . \tag{10.2}$$

Proof:

In case *A* we have $\frac{d}{d\alpha}(G_2(\alpha\mathbf{E}) - \gamma G_1(\alpha\mathbf{E})) = 0$. This expands to (10.1), when we substitute for the constant γ its value of $\frac{G_2(\mathbf{E})}{G_1(\mathbf{E})}$. A similar argument produces *B*, expanding the equation $\frac{d}{d\alpha}(G_2(E^1 + \alpha\theta_1, E^2 + \alpha\theta_2) - \gamma G_1(E^1 + \alpha\theta_1, E^2 + \alpha\theta_2)) = 0$. ■

We next define the conception of a *proportional allocation* for these economies.

Definition 10.2 An allocation is *proportional* if for all players *i*, $x_i = \frac{G_{l(i)}(\mathbf{E})E_i^{l(i)}}{\sum_{j=1}^{l} G_j(\mathbf{E})E^j} G(\mathbf{E})$.

That is, each player receives a share of output equal to the share of the *value* of his effort in the total value of effort, where each effort is evaluated at its marginal-product wage. An allocation is a *proportional solution* if it is proportional and Pareto efficient. This definition is taken from Roemer and Silvestre (1993).

Definition 10.3 An allocation is *generalized equal-division* if for all players *i*, $x_i = \frac{G_{l(i)}(\mathbf{E})}{\sum_{j=1}^{l} \lambda^j G_j(\mathbf{E})} G(\mathbf{E})$, where $\lambda^j = \#\{i \mid l(i) = j\}$.

Note that for $l=1$, proportional allocations are indeed the proportional allocations of Chapter 4, and likewise, the generalized equal-division allocation is an equal-division allocation of the earlier type. Note also that the sum over all players of the output shares in definition 10.3 is one.

To define Kantian allocations, we first write down the game that is induced in these economies when the allocation rule is the proportional rule. The payoff function for player *i* is given by:

$$V^i(E_1^{l(1)},E_2^{l(2)},...,E_n^{l(n)}) = u^i(\frac{G_{l(i)}(\mathbf{E})E_i^{l(i)}}{\sum_j G_j(\mathbf{E})E^j}G(\mathbf{E}),E_i^{l(i)}) \ . \qquad (10.3)$$

Definition 10.4 An effort vector $(E_1^{l(1)},E_2^{l(2)},...,E_n^{l(n)})$ is a *multiplicative Kantian equilibrium* ($K^{\times}$) for the game $\{V^i\}$ defined by (10.3) if:

$$(\forall i=1,...,n)(\arg\max_{r\geq 0} V^i(rE_1^{l(i)},...,rE_n^{l(n)})=1) \ . \qquad (10.4)$$

Now consider economies that allocate output according to generalized equal-division. The induced payoff functions of the game are:

$$\tilde{V}^i(E_1^{l(1)},...,E_n^{l(n)}) = u^i(\frac{G_{l(i)}(\mathbf{E})}{\sum_{j=1}^{l}\lambda^j G_j(\mathbf{E})}G(\mathbf{E}),E_i^{l(i)}) \ . \qquad (10.5)$$

Definition 10.5 An effort vector $(E_1^{l(1)},E_2^{l(2)},...,E_n^{l(n)})$ is an *additive Kantian equilibrium* (K^{+}) for the game $\{\tilde{V}^i\}$ defined by (10.5) if:

$$(\forall i=1,...,n)(\arg\max_{r\geq -\min_i E_i^{l(i)}} \tilde{V}^i(E_1^{l(i)}+r,...,E_n^{l(n)}+r)=0) \ . \qquad (10.6)$$

We can now state the main result:

Proposition 10.2

A. *Let G be homothetic. Let* $(E_1^{l(1)},E_2^{l(2)},...,E_n^{l(n)})$ *be an effort allocation such that* $E_i^{l(i)}>0$ *for all i, that is a multiplicative Kantian equilibrium for the game* $\{V^i\}$. *Then the induced allocation is Pareto efficient in the economy* $(u^1,...,u^n,G)$.

B. Let G be C-homothetic along the expansion path (λ^1,λ^2) . *Let* $(E_1^{l(1)},E_2^{l(2)},...,E_n^{l(n)})$ *be an additive Kantian equilibrium for the game* $\{\tilde{V}^i\}$. *Then the induced allocation is Pareto efficient in the economy* $(u^1,...,u^n,G)$.

Thus, under different conceptions of homotheticity on production, if each worker supplies only one kind of labor, then positive multiplicative Kantian equilibria are efficient in proportional economies, and any additive Kantian equilibrium is efficient in an equal-division economy.

The hypothesis on the production function in Part B seems very restrictive: G must be *C*-homothetic on an expansion path that depends upon the skills of the workers – since λ^j is the number of workers capable of supplying labor of type *j*. Let us apply this condition to the Chipman production function stated above, which can be written:

$$G(E^1,E^2)=1-\exp(-\lambda^2\lambda^1\bar{E}^1)-\exp(-\lambda^1\lambda^2\bar{E}^2)\ ,$$

where $\bar{E}^j=\dfrac{E^j}{\lambda^j}$, which is the average labor supplied by workers who supply labor of type *j*. The simplest example would be a production process where each worker supplies his own unique kind of labor: then the values of λ^j are both one, and the expansion path along which the isoquants have constant slope is the 45^0 line.

Proof of Proposition 10.2: (we let $l=2$)

Part A. Let us suppose that worker *i* supplies labor of type 1. The condition defining $K^\times$ equilibrium is then:

$$(\forall i=1,...,n)\quad \left.\frac{d}{dr}\right|_{r=1} u^i\left(\frac{G_{l(i)}(r\mathbf{E})E_i^{l(i)}}{G_1(r\mathbf{E})E^1+G_2(r\mathbf{E})E^2}G(r\mathbf{E}),rE_i^1\right)=0\ , \qquad (10.7)$$

which expands, for a player *i* who supplies labor of type 1, to:

$$u_1^i\cdot\left(\begin{array}{l}\dfrac{G_1E_1^1}{G_1E^1+G_2E^2}\nabla G\cdot\mathbf{E}+\\ G\dfrac{(G_1E^1+G_2E^2)E_i^1((\nabla G_1)\cdot\mathbf{E})-E_i^1G_1(E^1(\nabla G_1\cdot\mathbf{E})+E^2(\nabla G_2\cdot\mathbf{E}))}{(G_1E^1+G_2E^2)^2}\end{array}\right)+u_2^iE_1^1=0\ , \qquad (10.8)$$

where G and its derivatives are evaluated at $\mathbf{E}$, and the gradient vector $\nabla G_j = (G_{j1}, G_{j2})$ for j = 1,2. (Thus, $\mathbf{H} = (\nabla G_1, \nabla G_2)$.) The sufficient condition for Pareto efficiency , because the solution is interior, is $u_i^1 G_1 + u_i^2 = 0$, which states that i's marginal rate of substitution equals her marginal productivity. Noting that $\nabla G \cdot \mathbf{E} \equiv G_1 E^1 + G_2 E^2$, and dividing (10.8) through by the positive number E_1^1 , we see that (10.8) reduces to the efficiency condition if the second term in the coefficient of u_i^1 is zero; that is, we need to show that:

$$(G_1 E^1 + G_2 E^2)((\nabla G_1) \cdot \mathbf{E}) = G_1 (E^1 (\nabla G_1 \cdot \mathbf{E}) + E^2 (\nabla G_2 \cdot \mathbf{E}),$$

which is equivalent to:

$$(G_1 E^1 + G_2 E^2)(G_{11} E^1 + G_{12} E^2) = E^1 G_1 (G_{11} E^1 + G_{12} E^2) + E^2 G_2 (G_{21} E^1 + G_{22} E^2).$$

(10.9)

The reader can check that condition (10.9) reduces to condition (10.1), which is true by the hypothesis that G is homothetic, and Proposition 10.1. This proves part *A*.

Part B.

The allocation is an additive Kantian equilibrium for the generalized equal-division economy if and only if:

$$(\forall i = 1,...,n)$$

$$\left.\frac{d}{dr}\right|_{r=0} u^i\left(\frac{G_{l(i)}(E^1 + \lambda^1 r, E^2 + \lambda^2 r)}{\lambda^1 G_1(E^1 + \lambda^1 r, E^2 + \lambda^2 r) + \lambda^2 G_2(E^1 + \lambda^1 r, E^2 + \lambda^2 r)} G(E^1 + \lambda^1 r, E^2 + \lambda^2 r), E_i^1 + r\right) = 0 \quad .$$

(10.10)

Verbally, condition (10.10) states that when each player considers the counterfactual 'add a constant r to everyone's effort,' the optimal constant she would choose is $r = 0$.

For a player i for whom $l(i) = 1$, this expands to:

$$u_1^i \cdot \left(\begin{array}{l} \dfrac{G_1}{\lambda^1 G_1 + \lambda^2 G_2}(\lambda^1 G_1 + \lambda^2 G_2) + \\ G \dfrac{(\lambda^1 G_1 + \lambda^2 G_2)(\lambda^1 G_{11} + \lambda^2 G_{12}) - G_1((\lambda^1)^2 G_{11} + 2\lambda^1 \lambda^2 G_{12} + (\lambda^2)^2 G_{22})}{(\lambda^1 G_1 + \lambda^2 G_2)^2} \end{array} \right) + u_i^2 = 0.$$

(10.11)

Again, the efficiency condition is $u_1^i G_1 + u_2^i = 0$ for worker *i*. Condition (10.11) reduces to the efficiency condition exactly when the second term in the coefficient of u_1^i is zero; i.e., when:

$$(\lambda^1 G_1 + \lambda^2 G_2)(\lambda^1 G_{11} + \lambda^2 G_{12}) = G_1((\lambda^1)^2 G_{11} + 2\lambda^1\lambda^2 G_{12} + (\lambda^2)^2 G_{22}). \qquad (10.12)$$

The reader can check that condition (10.12) is equivalent to condition:

$$(\lambda^1, \lambda^2)\mathbf{H}\begin{pmatrix} -G_2 \\ G_1 \end{pmatrix} = 0,$$

which is condition (10.2) for the expansion path (λ^1, λ^2) . By the hypothesis of *C*-homotheticity for this expansion path, Part B is proved. ■

I do not believe there is a generalization of proposition 10.2 to the case where individual workers supply several kinds of labor. Indeed, it is perhaps remarkable that when each worker supplies a unique type of labor, the Kantian counterfactuals that 'work' are the simple proportional rule and the simple additive rule. Of course, a worker's share of output labor in the definitions of the (generalized) proportional and generalized equal-division allocation rule takes into account the marginal productivities of different kinds of labor. Having defined the allocation rules in this way, no further adjustment is needed when agents imagine the Kantian counterfactuals.

Consider this application of Proposition 10.2A. Suppose capital is a factor of production: let it be the first factor so E^1 denotes the input of capital and $E^2, ..., E^n$ now denote the contributions of the various kinds of labor. Suppose there are some individuals who possess capital but no labor, and they, too, have utility functions $u^i(x, E)$ where the provision of capital by them to production involves a disutility (perhaps it reduces consumption). According to Prop. 10.2A , if *G* is homothetic, then a positive multiplicative Kantian equilibrium is Pareto efficient. In this case, $G_1(\mathbf{E})$ is the marginal product of capital. In this equilibrium, every individual, whether the provider of capital

or labor, receives a share of the product proportional to his contribution, evaluated at its marginal-product price. If *G* exhibits constant returns to scale, then this allocation is identical to the Walrasian equilibrium allocation – that is, each receives a share of the product *equal* to his contribution evaluated at marginal-product prices, because the sum of contributions evaluated at marginal-product prices exhausts the total product. But if *G* exhibits decreasing returns then the allocation is not Walrasian, for it allocates the entire product in proportion to contributions, so evaluated. Contrast this with the Arrow-Debreu model, in which pure profits, which exist at Walrasian equilibrium with strictly concave production, are allocated to shareholders, while providers of capital (and labor) each receive a share of the product *equal to* (rather than proportional to) their contributions (always evaluated at marginal-product prices).

Words: 801

Final Remarks

I have taken from Michael Tomasello the claim that cooperation is unique to humans among the five species of great ape, and I have taken it for granted that the huge accomplishments of our species, in transforming the world and in developing our capabilities, have been in large part due to cooperation. Economics as a science has neglected cooperation; it has focused upon the competitive behavior of our species' members. It is time to attempt to apply the same methods of abstraction that have been so successful in competitive economic theory to analyze how we cooperate.

As I have discussed, the issue has not been entirely neglected. Von Neumann and Morgenstern (1944) made the first attempt in modern times, with the theory of cooperative games. More recently, cooperation has been the focus of many behavioral economists, some of whose work I have discussed. Finally, there has been a recent burgeoning of sophisticated explanations of cooperation based upon versions of Nash equilibrium, by researchers whom one might not call behavioral economists.

I think that all these attempts are off-base, for reasons I have given. Concerning the more recent developments, I have two main objections: first, I find the attempt to justify cooperation as a kind of Nash equilibrium (that is, one in which players are optimizing in an autarkic way) wrong-headed. I do not deny that some participants in a cooperative venture reason in the Nash manner, but I do not think successful cooperative projects can succeed if this kind of thinking is universal. More natural to me is that many cooperators think in the Kantian manner, and this thought process is induced by recognition of the commonality of their situations.

My second objection to many of the contributions of behavioral economics is that the frequent move is to include exotic arguments such as fairness in preferences of agents. I object to this approach for two reasons: first, from an aesthetic viewpoint, I prefer to keep preferences simple, and not to argue that *every* deviation we see from competitive behavior must be explained by deviations of preferences from classical ones. After all, as I have been at pains to argue, we need both preferences and an optimization protocol to produce choices, and I believe we should exploit the second component more than we have. Secondly, many 'behavioral' explanations of deviations from competitive behavior

are ad hoc: one must know *a piori* what the cooperative action is in order to define the conception of fairness that enters into preferences. In common-pool resource problems, and generally in economic problems involving public goods and bads, it is not *a priori* obvious what the cooperative action is. The Kantian approach enables at least the analyst to compute the cooperative solution, just as the Nash protocol enables the analyst to compute the competitive solution.

The theory of cooperation I have proposed is limited in scope. For the most part, it requires that individuals have unidimensional strategies. This may be due to my own limited vision; or it may be due to the fact that the Kantian protocols require that there exist a clear conception of when actions taken by different individuals can be considered to be similar, of the same kind. This is clearest when agents are 'identical,' as in symmetric games; it is somewhat clear if strategy spaces are unidimensional. Perhaps the Kantian approach is itself a limitation, and others will think of entirely different ways of conceptualizing cooperative thinking.

The theory I have proposed distinguishes between cooperation and altruism. It is a strength of the theory that psychological altruism is unnecessary for cooperation, because, I believe that cooperation is more accessible to humans than altruism. I believe, as well, that cooperation may lead to altruism in some cases, but I have not discussed how this may occur, because I have nothing new to say about it.

My own quick and dirty summary of human history is that it's a story of the extension of the boundaries of cooperation to ever larger groups. Between these groups, competition remains primary, which in extreme cases becomes war. Not only do the groups within which cooperation occurs become larger over time, but within these groups cooperation becomes more pervasive. I need hardly mention the caveat that this development is not monotonic, nor do I claim that we necessarily will reach a globally cooperative nirvana. We could destroy ourselves, and take down many other species in the process. If I have a political purpose in writing this short book, it is to shift the focus of economic theory, by providing an alternative view to the ruling conception of *homo economicus*. If I succeed in doing so, even in some small way, that may help economic theory contribute more than it has to developing our cooperative capacities and projects.

Words: 773

References

Akerlof, G. 1982. "Labor contracts as partial gift exchange," *Quar. J. Econ.* 97, 543-569

Alger, I. and J. Weibull, 2013. " Homo moralis – preference evolution under incomplete information and assortative matching," *Econometrica* 81, 2269-2302

Andreoni, J. 1990. "Impure altruism and donations to public goods: A theory of warm-glow giving," *Econ. J.* 100, 464-477

Bergstrom, T. 1995. "On the evolution of altruistic ethical rules for siblings," *Amer. Econ. Rev.* 85, 58-81

Boehm, S. 2012. *Moral origins,* New York: Basic Books

Bowles, S. and H. Gintis, 2004. *Microeconomics*, New York: Russell Sage Foundation

Bowles, S. and H. Gintis, 2011. *A cooperative species: Human reciprocity and its evolution,* Princeton University Press

Boyd, R., H. Gintis, S. Bowles and P. Richerson, 2003. "The evolution of altruistic punishment," *PNAS* 100, 3531-3535

Brekke, K.A., S. Kverndokk, and K. Nyborg, 2003. "An economic model of moral motivation," *J. Pub. Econ.* 87, 1967-1983

Chambers, C.P. and J.D. Moreno-Ternero, 2015. "Taxation and poverty," *Social Choice and Welfare*

Chipman, J. 1965. "A survey of the theory of international trade: Part 2, The neo-classical theory," *Econometrica* 33, 685-760

Cox, J. , E. Ostrom, J. Walker, A. Castillo, E. Coleman, R. Holahan, M. Schoon and B. Steed, 2009. "Trust in private and common property experiments," *Southern Econ. J.* 75, 957-975

Dufwenberg, M., P. Heidues, G. Kirchsteiger, F. Riedel and J. Sobel, 2011. "Other-Regarding Preferences in General Equilibrium," *Review of Economic Studies*,

Feddersen, T. 2004. "Rational choice theory and the paradox of not voting," *J. Econ. Persp.* 18, 99-112

Words: 773

Fehr, E. and H. Gintis, 2007. "Human motivation and social cooperation: Experimental and analytical foundations," *Annnual Review of Sociology* 33, 43-64

Fehr, E. and K.M. Schmidt, 1999. "A theory of fairness, competition and cooperation," *Quar. J. Econ.* 114, 817-868

Frey, B. and B. Torgler, 2007. "Tax morale and conditional cooperation," *J. Comparative Econ.* 35, 136-159

Gambetta, D. 2015. "What makes people tip," in C. Lopéz-Guerra and J. Maskiver (eds.), *Rationality, democracy and justice: The legacy of Jon Elster,* 97-114, Cambridge University Press

Gintis, H. 2000. "Strong reciprocity and human sociality," *J . Theor. Biology* 206, 169-179

Hardin, G. 1968. "The tragedy of the commons," *Science* 162, 1243-1248

Kobayashi, H. and S. Kohshima, 2001. "Unique morphology of the human eye and its adaptive meaning: Comparative studies on external morphology of the primate eye," *J. Human Evol.* 52, 314-320

Ju, B-G., E. Miyagama and T. Sakai, 2007. "Non-manipulable division rules in claims problems and generalizations," *J. Econ. Theory* 132, 1-26

Laffont, J-J. 1975. "Macroeconomic constraints, economic efficiency, and ethics: An introduction to Kantian economics," *Economica* 42, 430-437

Makowski, L. and J. Ostroy, 2001. "Perfect competition and the creativity of the market," *J. Econ. Lit.* 39, 479-535

Mas-Colell, A. 1987. "Cooperative equilibrium," in J. Eatwell, M. Milgate and P. Newman, *The Palgrave dictionary of economics,* vol. 1, London: Macmillan

Mas-Colell, A. and J. Silvestre, 1989. "Cost share equilibria: A Lindahlian approach," *J. Econ. Theory* 47, 239-256

Moulin, H. 1987. "Equal or proportional division of a surplus, and other methods," *International J. Game Theory* 16, 161-186

Olson, M. 1965. *The logic of collective action* , Harvard University Press

Rabin, M. 2003. "Incorporating fairness into game theory and economics," in C. Camerer, G. Loewenstein and M. Rabin, *Advances in behavioral economics,* chapter 10, New York : Russell Sage Foundation and Princeton: Princeton University Press

Words: 773

Rabin, M. 2004. "Incorporating fairness into game theory and economics," in C. Camerer, G. Loewenstein and M. Rabin, *Advances in behavioral economics,* Russell Sage Foundation and Princeton University Press

Richter, A. and J. Grasman, 2013. "The transmission of sustainable harvesting norms when agents are traditionally cooperative," *Ecological Econ.* 93, 202-209

Roemer, J. 1996. *Theories of distributive justice,* Harvard University Press

Roemer, J. 2006. "Party competition under private and public financing: A comparison nof institutions," *Advances in theoretical economics* 6, Issue 1, article 2, http://www.bepress.com/bejte/advances/vol6/iss1/art2

Roemer, J. and J. Silvestre, 1993. "The proportional solution for economies with both private and public ownership," *J. Econ. Theory* 59, 426-444

Sugden, R. 1982. "On the economics of philanthropy," *Econ. J.* 92, 341-350

Thomson, W. 2015. "For claims problems, compromising between the proportional and constrained equal awards rules," *Econ. Theory*

Tomasello, M. 2014a. *A natural history of human thinking,* Cambridge MA: Harvard University Press

Tomasello, M. 2014b. "The ultra-social animal," *European J. Social Psych.* 2014

Von Neumann, J. and O. Morgenstern, 1944. *Theory of games and economic behavior*, Princeton: Princeton University Press

Walker, J. and E. Ostrom, 2009. "Trust and reciprocity as foundations for cooperation," in K. Cook, M. Levi and R. Hardin, *Whom can we trust?* Chapter 4, New York: Russell Sage Foundation