KANTIAN ALLOCATIONS

By

John E. Roemer

September 2006

COWLES FOUNDATION DISCUSSION PAPER NO. 1582

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

August 21, 2006

"Kantian allocations"

by

John E. Roemer*
Departments of Political Science and Economics
Yale University
PO Box 208301
New Haven CT 06520

john.roemer@yale.edu
phone: 203-432-5249
fax:203-432-6196

* I am grateful to Joaquim Silvestre for discussions of the topic, and to Juan Moreno-Terrnero for correcting an error in an earlier version. I thank Colin Stewart and R.E. Howe for their mathematical contributions, noted in the text.

Abstract. Several authors in the economics literature have referred to Kantian behavior, informally, as a kind of cooperation. We model this notion precisely, and define two kinds of Kantian allocation. An set of strategies by players is *Kantian* if, informally, *no* player would advocate that *all* players change their strategies in the 'same kind of way.' We prove existence and Pareto efficiency of Kantian allocations. The proportional solution in a production economy with a common access technology emerges as a special case. We study whether Kantian behavior can 'resolve' the prisoners' dilemma and the voting paradox. It turns out that Kant's categorical imperative only implies cooperation (solidaristic behavior) conditional upon the rewards to cooperation being sufficiently great, perhaps a sobering thought for philosophical Kantians who believe that Kant's categorical imperative implies a strong kind of solidarity.

Key words: cooperative solution, proportional solution, voting paradox, prisoners' dilemma, Kant, categorical imperative

JEL codes: C71, D63

1. Cooperation and inefficiency

Kant's categorical imperative says that one should take those actions and only those actions which one would like all others to take as well. The categorical imperative is a cooperative norm. Our purpose in this article is to formulate this idea in a precise manner, on a class of economic environments, and to remark upon the properties of a 'Kantian equilibrium:' in particular, its existence and efficiency properties, and its relationship to cooperative behavior in several examples.

There is a history to this idea in the economics literature. Laffont (1977) writes:

> To give substance to the concept of a new ethics, we postulate that a typical agent assumes (according to Kant's morals) that the other agents will act as he does, and he maximizes his utility function under this new constraint…. Our proposition is then equivalent to a special assumption of others' behaviour. It is clear that the meaning of 'the same action' will depend on the model and will mean usually 'the same *kind* of action.'

In our formal definition to follow, we will define what taking 'the same *kind* of action' means in a precise way.

Sugden (1982) discusses philanthropy, and argues, with empirical evidence, that 'the Nash assumption,' that donors take the contributions of others as given, is not empirically verified. He writes: "Or suppose that each person, instead of having Nash conjectures, believes that if he gives a certain minimum sum of money, everyone else will do the same, but that if he gives less, everyone else will give nothing." This is his Kantian premise.

Feddersen (2004) offers a 'group-based ethical model' to explain the voting paradox. He writes "First, ethical agents evaluate alternative behavioral rules in a

Kantian manner by comparing the outcomes that would occur if everyone who shares their preferences were to act according to the same rule."

Finally, I introduced the definition of Kantian allocation formally in Roemer (1996, chapter 6). What I present below is a considerable generalization of what I presented in that book.

To motivate the definition, consider the problem associated with the tragedy of the commons. A set of fishers must expend labor on a lake to catch fish, but there is a congestion problem, so that the fish caught per unit of labor decreases with the number of total hours expended in fishing. Each fisher has a utility function over fish caught and labor expended. In the Nash equilibrium of the game where each fisher's strategy is his labor choice, there is overfishing: the equilibrium is Pareto inefficient, and all would profit from a small decrease in their labor expenditures. Some kind of cooperation is necessary to solve the problem.

Consider this more abstract problem. There is a set of *n* agents, each with a utility function $u^i : \mathbb{R}^n_+ \to \mathbb{R}$. Denote the argument of the utility function by $(L^1,...,L^n)$, to be thought of as a vector of efforts expended by the *n* agents. Define $L^{-i} = (L^1,...,L^{i-1},L^{i+1},...,L^n)$. Suppose that u^i is strictly monotone decreasing[1] in L^{-i} for each *i*. This set-up describes the fisher problem, where $u^i(L)$ is the utility of fisher *i* if the vector of labors expended by all fishers is *L*. The more the others fish, the less efficient will *i*'s labor be on the lake. We put no restriction, for the moment, on the derivative of u^i with respect to L^i. In the fisher problem, an appropriate assumption would be that for any vector L^{-i}, $u^i(L^i,L^{-i})$ is increasing at $L^i = 0$ -- meaning that the

[1] That is, u^i is strictly decreasing in every component of L^{-i}.

fisher wants to avoid at all costs zero fish consumption – and that for any vector L^{-i}, for sufficiently large L^i, $u^i(L^i, L^{-i})$ is decreasing in i' s effort – meaning that the disutility of labor eventually becomes paramount.

The fisher problem is one where other agents' efforts create a public bad. The opposite problem is when the efforts of the community create a public good. We can also represent this problem by a profile of functions $\{u^i \mid i = 1,...,n\}$ defined on the domain $\mathbb{R}^n_+$; this time, however, the appropriate assumption is that u^i is an increasing function in L^{-i}.

We will say that the profile of utility functions $\{u^i\}$ is *monotonic* if

$(\forall i)(u^i$ is strictly monotone increasing in $L^{-i})$, or

$(\forall i)(u^i$ is strictly monotone decreasing in $L^{-i})$

We formalize the Kantian notion as follows.

Definition 1 Let $\{u^i \mid i = 1,...,n\}$ be a monotonic profile. An allocation of effort $\hat{L} = (\hat{L}^1,...,\hat{L}^n)$ with the property that *no agent* would prefer that *all agents* change their efforts by the same factor -- in other words:

$$(\forall i = 1,...,m)(\forall a \geq 0)(u^i(\hat{L}) \geq u^i(a\hat{L})) \tag{1.1}$$

is a *Kantian allocation*.

One might think of Kantian allocations as ones that are stable with respect to some set of *focal alternatives*. At an allocation L, it might be salient to ask, "Why don't we all reduce our labor by 5%?" The proposal that we all do the 'same thing' is a focal point. Thus, Kantian solutions are stable with respect to this class of focal alternatives.

One might also view the Kantian concept as a model of 'magical thinking[2].' An agent thinks, "If I do not vote, neither will anyone like me vote; if I vote, then others like me will as well." There are two interpretations of that sentence. The magical one is that my behavior will *cause* a like behavior in people who are similar to me. A less magical interpretation is that I am a token – if, by whatever thought process, I come to choose a certain action, then all others in the class (of which I am a token) will come to the same decision. Consequently, the justification of the Kantian concept need not be ethical; it could be 'strategic' in this sense of magical thinking or tokenism.

2. Efficiency and existence

The results presented below apply to both cases discussed – where, for all *i*, u^i is strictly increasing in L^{-i}, or for all *i*, u^i is strictly decreasing in L^{-i}. For expositional simplicity, we will henceforth assume the first case, except in some examples.

Let us define a second conception of salient variation. We say an allocation $\hat{L}$ is *v-Kantian* for a vector $v \in \mathbb{R}^n_{++}$ if

$$(\forall i)(\forall a \in \mathbb{R})(a\hat{L} \geqq 0 \Rightarrow u^i(\hat{L}) \geq u^i(\hat{L} + av)). \tag{2.1}$$

The simplest case is when $v = (1,1,...,1)$. Then $\hat{L}$ is *v*-Kantian when no agent would advocate that all agents increase (or decrease) their effort by any given amount, *a*.

Thus, a Kantian allocation is one where the focal deviation is multiplicative with respect to effort contributions, and a *v*-Kantian allocation is one where the focal deviation is additive.

---

[2] For a discussion of magical thinking as an explanation of behavior, see Elster (1989).

In a Nash allocation, all agents contemplate a deviation in their effort, while assuming that all others' efforts remain fixed. Nash allocations are in general Pareto inefficient. One sense in which Kantian allocations can be viewed as cooperative is that they are efficient. We have:

Theorem 1. Let the allocation $\hat{L} \in \mathbb{R}^n_+, \hat{L} \neq 0$ be either Kantian or v-Kantian, for some $v \in \mathbb{R}^n_{++}$. Then $\hat{L}$ is Pareto efficient.

Proof[3]:

1. Let $\hat{L}$ be Kantian. Suppose, to the contrary, that $\hat{L}$ is not Pareto efficient; let L be an allocation such that, for all i, $u^i(L) \geq u^i(\hat{L})$, with strict inequality for at least one i. Because $\hat{L}$ is Kantian, it follows that $\hat{L}$ is a positive vector: for if its jth component were zero, then j would advocate multiplying everyone's labor by a large positive number, because $\hat{L}$ is, by hypothesis, positive in at least one component, and u^j is strictly increasing in L^{-j}. Therefore, we can define

$$a^* = \max_i \frac{L^i}{\hat{L}^i}. \tag{2.2}$$

Let this maximum be assumed at the index i^*. If (2.2) holds for *all* indices i, then we have a direct contradiction to the assumption that $\hat{L}$ is Kantian. So we may assume that the vector $a^* \hat{L} \geq L$ with equality holding in the i^* component, but strict inequality holding in at least one component. Note now that $a^* \neq 1$: for otherwise, $u^{i^*}(L) < u^{i^*}(\hat{L})$, by invoking monotonicity. It follows that

$$u^{i^*}(a^* \hat{L}) > u^{i^*}(L) \geq u^{i^*}(\hat{L}). \tag{2.3}$$

[3] This generalizes a proof due to Colin Stewart.

The first inequality is true because u^{i^*} is monotone in L^{-i^*}. But (2.3) contradicts the assumption that $\hat{L}$ is Kantian: i^* would like to alter all efforts by the factor a^*. This contradiction proves the claim.

2. Assume the second case, that $\hat{L}$ is v-Kantian[4]. Suppose $\hat{L}$ is Pareto-dominated by the allocation L. Define

$$a^* = \max_i \frac{L^i - \hat{L}^i}{v^i}, \qquad (2.4)$$

which is well-defined since v is a positive vector. Suppose the maximum is assumed at the index i^*. As above, it immediately follows that the maximum is not assumed for all indices – or else $\hat{L}$ would not be v-Kantian. Now we have

$$u^{i^*}(\hat{L} + a^* v) > u^{i^*}(L) \geq u^{i^*}(\hat{L}), \qquad (2.5)$$

contradicting the fact that $\hat{L}$ is v-Kantian. ■

Theorem 1 is striking. The fact that an allocation $\hat{L}$ is Kantian means that it (weakly) Pareto-dominates all allocations on the ray through $\hat{L}$: this implies that it is Pareto efficient! Likewise, if $\hat{L}$ is v-Kantian, it is unanimously preferred to all other allocations on a unidimensional 'subspace' containing $\hat{L}$; but this implies that it is efficient.

We now address existence.

Assumption A. Let $\{u^i\}$ be an environment such that:

[4] This part of the proof need not assume that $\hat{L} \neq 0$.

(a) for every *i*, there is a number $\varepsilon^i > 0$ such that, for all vectors L^{-i}, $u^i(L^i, L^{-i})$ is strictly increasing in L^i, for $L^i < \varepsilon^i$.

(b) for every *i*, there is a number γ^i, such that for all vectors $L^{-i}, u^i(L^i, L^{-i})$ is strictly decreasing in L^i, for $L^i > \gamma^i$.

Consider the fishing village described in section 1. Assumption A is reasonable for it. Part (a) says that, no matter what the congestion on the lake, each fisher *i* must consume some fish – so his utility increases as the fisher increases his or her labor from zero to some positive number. Part (b) says that eventually expending more labor becomes a disutility, even if there is very little congestion on the lake. Assumption A is also reasonable for a public-good problem. Suppose that the level of public good is $g(L^1,...,L^N)$ where L^i is the contribution of agent *i*. Suppose utility functions are given by $u^i(L) = v^i(g(L)) - c^i(L^i)$, where c^i is the cost-of-effort function for agent *i*. If $\frac{dc^i}{dL^i}(0) = 0$, and *g* and *v* are strictly increasing, then assumption A(a) holds. Of course it is normal that A(b) would hold.

<u>Theorem</u> <u>2</u>. Let assumption *A* hold for the environment $\{u^i\}$. Then:

(a) a Kantian allocation exists;

(b) for any $v \in \mathbb{R}^n_{++}$, a *v*-Kantian allocation exists.

<u>Proof</u> <u>of</u> (a):

1. The set $\Lambda = \{(L^1,...,L^N) \in \mathbb{R}^n \mid \varepsilon^i \le L^i \le \gamma^i\}$ is convex and compact.

2. For any $L \in \Lambda$, define:

$$\alpha^i(L) = \arg\max_a u^i(aL). \tag{2.6}$$

By assumption *A,* these numbers are well-defined. Furthermore, note that

$$\varepsilon^i \le \alpha^i(L)L^i \le \gamma^i \,, \tag{2.7}$$

a fact which follows from assumption *A* and the monotoncity of u^i in L^{-i}. Note also that the functions $\alpha^i(\cdot)$ are continuous on Λ by the Berge maximum theorem.

3. Define the function $\Phi:\Lambda \to \Lambda$ by

$$\Phi(L) = (\alpha^1(L)L^1,...,\alpha^N(L)L^N). \tag{2.8}$$

We have shown that Φ indeed has its range as Λ. Consequently, it is a continuous function mapping a compact, convex set into itself, and therefore possesses a fixed point, which we denote $\hat{L}$.

4. But $\Phi(\hat{L}) = \hat{L}$ implies that for all *i,* $\alpha^i(\hat{L}) = 1$. By definition of the functions α^i, it follows immediately that $\hat{L}$ is a Kantian allocation.

<u>Proof</u> <u>of</u> (b):

5. Define the functions:

$$\beta^i(L) = \arg\max_a u^i(L + av). \tag{2.9}$$

Exactly the same argument as above shows that the function $\mathrm{B}:\Lambda \to \Lambda$ defined by

$$B(L) = (L^1 + \beta^1(L)v^1,...,L^N + \beta^N(L)v^N) \tag{2.10}$$

possesses a fixed point, which is, indeed, a *v*-Kantian allocation. ■

We conclude this section by presenting a converse of theorem 1. When can an arbitrary Pareto –efficient allocation for an environment $\{u^i\}$ be rationalized as *v*-Kantian, for some choice of *v*? Perhaps the most salient choice of *v* is the vector of ones: $v = (1,1,..,1)$. As we have remarked, if an allocation is $(1,1...,1)$-Kantian, then nobody would prefer that everyone increase (decrease) his effort by the same absolute amount. But there might be other salient choices of v. For instance, 'effort' might be monetary contribution to a public project, and we might choose v^i proportional to i' s income or wealth.

It turns out that essentially all Pareto efficient allocations are v-Kantian, at least with some regularity restrictions. Let $\{u^i\}$ be an environment where the u^i are differentiable and concave. Let $\hat{L}$ be an interior Pareto efficient allocation. Let $u^i(\hat{L}) = k^i$. Then Pareto efficiency implies that $\hat{L}$ solves the program:

$$\begin{aligned}&\max u^1(L)\\&s.t.\\&u^j(L) \geq k^j, j = 2,...,N\end{aligned}$$

and all the constraints are binding. By the Kuhn-Tucker theorem, there are non-negative numbers $\lambda^2,...,\lambda^N$ such that

$$\nabla u^1(\hat{L}) + \sum_{j=2}^{N} \lambda^j \nabla u^j(\hat{L}) = 0. \tag{2.11}$$

Call the allocation $\hat{L}$ *regular* if $\lambda^j > 0, j = 2,...,N.$

We have:

<u>Theorem</u> <u>3</u>. Let $\{u^i\}$ be a concave, differentiable environment, and let $\hat{L}$ be an interior, regular Pareto efficient allocation. Then there exists $v \in \mathbb{R}^n_{++}$ such that $\hat{L}$ is v-Kantian.

To prove the theorem, we invoke a linear -algebraic fact:

Lemma 1 Let T be an $n \times n$ matrix whose off-diagonal elements are positive. Suppose there exists a positive (row) vector $y \in \mathbb{R}^n$ such that yT=0. Then there exists a positive (column) vector v such that Tv=0.

Proof: (thanks to R. E. Howe)

1. Let I denote the $n \times n$ identity matrix. We may choose a positive number c such that $B = T + cI$ is a positive matrix. B has a unique positive eigenvector, up to scale, call it x, associated with its maximum eignevalue, λ. Thus $xB = \lambda x$. But $yB = yT + cy = cy$, so y is a positive eigenvector of B, and hence c=λ. Consequently the transpose of B possesses a positive eigenvector with eigenvalue c, call it v: thus, $Bv = Tv + cv = cv$, and so $Tv = 0$, as required. ■

Proof of theorem 3:

1. Define the n x n matrix $T = \begin{pmatrix} \nabla u^1(\hat{L}) \\ \ldots \\ \nabla u^n(\hat{L}) \end{pmatrix}$. Define the vector y=$(1, \lambda^2, \ldots, \lambda^N)$ from (3.1), which is a positive vector. Eqn. (3.1) says that yT=0.

Note that the off-diagonal elements of T are positive, by monotonicity. It follows by the lemma that a positive vector v exists such that Tv=0.

2. Recall that an allocation L is z-Kantian for a vector z just in case

$$\forall i \quad 0 = \arg\max_a u^i(L + az). \tag{2.12}$$

In the differentiable case, the first-order condition for these statements is:

$$\forall i \quad \nabla u^i(L) \cdot z = 0. \tag{2.13}$$

But if the u^i are concave, then (3.2) is also a sufficient condition (3.2) to hold. Note that (3.3) just says Tz=0. It follows from step 1 that L is v-Kantian, because Tv=0. ■

Theorem 3 may be somewhat of a curiosum. One is left asking, 'How does one justify an arbitrary positive vector v as the 'social norm' for defining what it means for agents to 'vary in the same way' from a given allocation?' Nevertheless, I present it, because it offers a potential ethical justification for any Pareto efficient allocation: one must justify the social norm implied by its associated vector v, which renders the allocation v-Kantian. This might indeed offer a way of refining the set of Pareto efficient allocations to ones that are potentially desirable. If L is an efficient vector of monetary contributions to a public project, one might say that a necessary condition for L to be desirable is that, if v is the vector with respect to which L is v-Kantian, then i richer than j should imply $v^i \geq v^j$. This would mean that, in considering acceptable variations from L, rich people are expected to increase their contributions by at least as much as poor people.

Theorem 3 tells us that we should not expect to find any interesting characterization of v-Kantian allocations – because the concept is not restrictive, beyond implying Pareto efficiency.

3. The proportional solution

Consider the following economy with production, a formalization of the fishing village discussed earlier. Let {1,…,N} be a society of N agents, where agent i has a utility function $v^i(x^i, L^i)$ over output and labor. Agent i has skill level s^i. The amount of efficiency labor expended if the vector of efforts is $L=(L^1,\dots,L^N)$ is $\ell=\sum s^i L^i$. Let $f(\ell)$ be the total output if ℓ units of efficiency labor are expended; let f be strictly concave.

Definition 2 A *common access allocation* is an allocation

$$\{(x^i,L^i)_{i=1,\dots,N} \mid \sum x^i = f(\sum s^i L^i) \text{ and } x^i = \frac{s^i L^i}{\sum s^j L^j}\sum x^j\}.$$

The first property of a common access allocation says that it is feasible; the second property says that output is distributed in proportion to the efficiency units of labor applied by individuals. The second property is meant to model production on a common-access resource where producers work individually, such as a lake (where fishers catch fish) or a forest (where woodcutters harvest trees). It is, of course, an idea that abstracts away from random variations in output. Absent those variations – think of a lake in which fish are distributed perfectly homogeneously – the amount of output a given producer will acquire is just proportional to the (efficiency units of) labor he expends.

Roemer and Silvestre (1993) defined:

Definition 3 A *proportional solution* for the production economy (v,s,f) where v is the vector of utility functions and s is the skill vector, is a common access allocation which is Pareto efficient.

The 'tragedy of the commons' is a colorful description of the Nash equilibrium of the economy (v,s,f). In the game, each agent's strategy is his labor supply. A Nash equilibrium is a strategy vector with the property that, for every *i*, given L^{-i}, agent *i*'s effort level solves the problem:

$$\max_L v^i\left(\frac{s^i L}{\sum_{j\neq i} s^j L^j + s^i L} f\left(\sum_{j\neq i} s^j L^j + s^i L\right), L\right).$$

In other words, given the labor contributions of all others, agent *i*'s contribution must maximize his utility over all common access allocations that he can induce by choice of his own effort level. It is well-known that, because *f* is strictly concave, the Nash equilibria, so defined, are inefficient.

A proportional solution is thus a resolution of the tragedy of the commons. It is, if you will, an implementation of the socialist dictum "to each according to his labor", *plus* Pareto efficiency.

Denote $\ell = \sum s^i L^i$. Define the utility functions:

$$u^i(L^1,...,L^N) = v^i\left(\frac{s^i L^i}{\ell} f(\ell), L^i\right). \tag{3.1}$$

Note that u^i is, indeed, strictly monotone decreasing in L^{-i} by the strict concavity of *f*. It follows from Theorem 1 that a (non-zero) Kantian allocation for the environment $\{u^i\}$ is a proportional solution for the village (v,s,f).

We have the following:

Corollary Let *(v,s,f)* be a production economy such that for all *i*, $\frac{\partial v^i(0,0)}{\partial x} = \infty$ and for large L and *x*>0, $\frac{\partial v^i(x,L)}{\partial L} = -\infty$. Then a proportional solution exists (and it is a Kantian allocation).

Proof:

Define the functions $\{u^i\}$ as in (3.1). This family inherits the property of assumption A from the assumptions in the premise of the corollary. It follows from theorem 2 that a Kantian solution exists for this environment. But this is a proportional solution for the production economy. ■

This result is to be contrasted with the (rather difficult) proof of the existence of a proportional solution in Roemer and Silvestre (1993)[5]. The earlier result required concavity assumptions on the utility functions v. Note that the corollary requires no such assumptions. Nevertheless, the Roemer-Silvestre existence theorem will apply in some cases where the premise of the corollary does not hold.

The proportional solution embodies a simple concept of fairness – plus Pareto efficiency. Here, then, we see the sense in which Kantian allocations have an ethical aspect. Indeed, it can be shown that, in the production economy, the converse holds as well: every Kantian allocation for the production economy (v,s,f) is a proportional solution (see Roemer [1996, Theorem 6.6]).

4. The limited cooperation of Kantian allocations

We have emphasized, to this point, the cooperative nature of the Kantian concept. In this section, we show, by means of two examples, that the kind of cooperation associated with the categorical imperative, as here modeled, is *limited* in nature.

A. The prisoners' dilemma

[5] It must be added that the proportional solution in Roemer and Silvestre (1993) is defined on a more general space of economies than the definition offered here.

Sometimes, efforts are chosen from a discrete set: for example, {Vote, Don't Vote}, or in the prisoners' dilemma, {Cooperate, Defect}. To adapt the notion of Kantian allocation to these environments, we will consider mixed strategies. The strategy for each agent is then a probability distribution over a common set of actions.

Consider the prisoners' dilemma, where the payoffs, in terms of von Neumann-Morgenstern utilities of the two players, are given by:

	Cooperate	Defect
Cooperate	*(b,b)*	*(d,a)*
Defect	*(a,d)*	*(c,c)*

where $a > b > c > d$. The strategy space for both players is [0,1], where a strategy of *p* means 'cooperate with probability *p,* defect with probability $1 - p$.'

We will examine the symmetric strategy profiles where both players play the same strategy. At a strategy pair (*p,p*) the expected utility of each player is given by:

$$V(p,p) = p^2 b + (1-p)^2 c + p(1-p)(a+d).$$

Define the function

$$\begin{aligned}\varphi(\alpha) &= V(\alpha p, \alpha p) = \alpha^2 p^2 b + (1-\alpha p)^2 c + \alpha p(1-\alpha p)(a+d) \\ &= \alpha^2 p^2 (b + c - a - d) + \alpha p(a + d - 2c) + c\end{aligned}$$

Then *(p,p*) is a *symmetric Kantian allocation* if and only if φ is maximized at 1 on the interval $[0,1/p]$.

<u>Theorem</u> <u>4</u> (*p,p*) is a symmetric Kantian allocation for the PD game iff

(a) *p*= 0, or

(b) $b \geq \frac{a+d}{2}$ and $p = 1,$ or

(c) $b \leq \frac{a+d}{2}$ and $p = \frac{a+d-2c}{2(a+d-b-c)}$.

Part (a) of the theorem is of little interest – (0,0) is a symmetric Kantian allocation for the same reason that $\hat{L} = 0$ is a Kantian allocation in the environment of theorem 1. What is interesting is that full cooperation – that is, playing *p*=1 – is only a symmetric Kantian allocation if the returns to cooperation (*b*) are sufficiently large. Otherwise (part (c)), cooperation is limited in extent (i.e., *p*<1).

<u>Proof of theorem 4</u>:

1. Obviously *p*=0 is Kantian: neither player can gain by calling for expansion of the probabilities by any factor.

2. For positive *p*, note that φ is a convex function if $b + c > a + d$ and is a concave function if $b + c < a + d$.

Case (i). *p*>0 and $b + c > a + d$

(*p,p*) is Kantian just in case φ is maximized at $\alpha = 1$. But in the convex case, φ is maximized at either 0 or 1/*p*. Thus, (*p,p*) is Kantian if and only if *p*=1 and $\varphi(1) \geq \varphi(0)$. But this last inequality says $b \geq c$, which is true. Therefore (1,1) is the unique Kantian allocation in this case with positive *p*. Note that, in this case, $b > \frac{a+d}{2}$, because *b*>c.

Case (ii) *p*>0 and $b + c < a + d$.

Since φ is concave, (*p,p*) is Kantian if and only if

(a) either *p*=1 and $\frac{d\varphi}{d\alpha}(1) \geq 0$, or

(b) $p<1$ and $\frac{d\varphi}{d\alpha}(1)=0$.

Calculate that $\frac{d\varphi}{d\alpha}(1)=0$ if an only if $p = p^* = \frac{a+d-2c}{2(a+d-(b+c))}$. It is easy to check that $p^* \in (0,1)$ in this case, and so (p^*,p^*) is Kantian.

On the other hand, if p=1, then $\frac{d\varphi}{d\alpha}(1)=2(b+c)-(a+d)-2c$ which is non-negative if and only if $b \geq \frac{a+d}{2}$. This proves part (c) and the result. ■

I believe this a sobering result: Kantianism is a less solidaristic concept than many might have thought. John Rawls (1971), for instance, described himself as a Kantian, and his 'difference principle' is considerably more solidaristic than the Kantian behavior described in theorem 4.

B. The voting paradox

Suppose there are two policies, which we denote α and β. The electorate is split into two groups: those who favor α and those who favor β. Denote these groups *A* and *B* and think of *A* and *B* as large populations. A general election is to be held to decide upon a policy, by majority vote. The strategy for an individual will be the probability *p* that she chooses to vote. Once each individual has chosen his/her strategy, the outcome of the election is a random variable, which is derived from the sum of the independently distributed random variables of the members of the two sides.

Each individual has von Neumann Morgenstern preferences over lotteries on four 'prizes.' The four prizes for an individual 'I' are {(α wins, I vote), (α wins, I don't vote), (β wins, I vote), (β wins, I don't vote)}. Assuming the cost of voting is independent of which policy wins, we can choose the vNM utility function for a member of group *A* to have the values $(1-b,1,-b,0)$ on these four prizes, respectively. That is, we normalize the value of the gain from one's favored policy at one, and the cost of voting at *b*. Thus, the individual's preferences are determined by the parameter *b*.

Suppose the probability of victory of α is some number π; then from the individual's viewpoint, the lottery associated with his choice of strategy over the four prizes associates to them the probabilities $(p\pi,(1-p)\pi,p(1-\pi),(1-p)(1-\pi))$. It follows that the expected utility of an individual in *A* is $\pi - pb$. In like manner, the expected utility of an individual in *B* is $1-\pi-pb$.

Suppose that every individual in *A* chooses the same probability *p* and every individual in *B* chooses the same probability *q*. Then we may write $\pi = \pi(p,q)$. We will derive the function $\pi(\cdot)$ below. But first we define our equilibrium concept. The idea is that each side behaves cooperatively in the Kantian sense, which is to say, that the strategy that members of each group choose comprise a Kantian equilibrium, given the choice of the other group. That is:

<u>Definition</u> <u>4</u> A *symmetric voting equilibrium* is a pair of probabilities (*p,q*) such that

(i) given *q*, *p* maximizes $\pi(p,q)-pb$

(ii) given *p*, *q* maximizes $1-\pi(p,q)-qb$.

Clearly, this is a special case of a more general concept of Kantian behavior where individuals within each coalition choose different probabilities. It seems reasonable to restrict examination to the simpler case of symmetric choice within each coalition.

In other words, the game is equivalent to a Nash game, where each coalition acts as a single co-ordinated player.

We have:

Lemma 2 Suppose that the behavior of each member of *A* is governed by the random variable

prob [vote] = *p*, prob [don't vote] = 1-*p*

and the behavior of each member of *B* is governed by the random variable

prob [vote] = *q*, prob [don't vote] = 1-*q*

and suppose that these random variables are independently distributed. Denote by *A* and *B* the sizes of the two coalitions. If *A* and *B* are large, then

$$\pi(p,q) \cong 1-\Phi(\frac{-\mu}{\sigma}),$$

where $\mu = \sqrt{A}(p - q\frac{B}{A})$, $\sigma = \sqrt{p(1-p) + \frac{B}{A}q(1-q)}$, and Φ is the CDF of the standard normal variate.

Proof:

Define the random variable

$$X = \begin{cases} 1 \text{ with prob } p \\ 0 \text{ with prob } 1-p \end{cases}.$$

The behavior of each individual in coalition *A* is i.i.d, described by *X*. The number of people of vote for α will be the realization of $\sum_{i \in A} X^i$.

2. The mean of *X* is *p* and its variance is $p(1-p)$. The random variable

$$\hat{X} = \frac{X-p}{\sqrt{p(1-p)}}$$

has a mean of zero and a variance of unity.

3. Hence, by the central limit theorem, because A is large, the distribution of $\sum_{i\in A}\frac{\hat{X}^i}{\sqrt{A}}$ is approximately standard normal.

4. In like manner, define the random variables

$$Y=\begin{cases}1 \text{ with prob } q\\ 0 \text{ with prob } 1-q\end{cases},\quad \hat{Y}=\frac{Y-q}{\sqrt{q(1-q)}}.$$

Then the distribution of $\sum_{j\in B}\frac{\hat{Y}^j}{\sqrt{B}}$ is approximately standard normal.

5. Now $\pi(p,q)=\Pr[\sum_{i\in A}X^i>\sum_{j\in B}Y^j]$. By the central limit theorem, it therefore follows that

$$\frac{\sum_{i\in A}X^i}{\sqrt{A}}\sim N[p\sqrt{A},\sqrt{p(1-p)}],$$

where $N[m,s]$ is the normal variate with mean m and standard deviation s, and

$$\frac{\sum_{j\in B}Y^j}{\sqrt{B}}\sim N[q\sqrt{B},\sqrt{q(1-q)}].$$

Now we calculate

$$\Pr[\sum_{i\in A}X^i>\sum_{j\in B}Y^j]=\Pr[\sum\frac{X^i}{\sqrt{A}}>\sqrt{\frac{B}{A}}\sum\frac{Y^j}{\sqrt{B}}]=$$
$$\Pr[N[p\sqrt{A},\sqrt{p(1-p)}]>N[\sqrt{\frac{B}{A}}q\sqrt{B},\sqrt{\frac{B}{A}}\sqrt{q(1-q)}]=$$
$$\Pr[N[\sqrt{A}(p-q\frac{B}{A}),\sqrt{p(1-p)+\frac{B}{A}q(1-q)}]>0],$$

where the last step uses the fact that the difference of two independent normal variates is a normal variate whose mean is the difference of the means of the first two, and whose variance is the sum of the variances of the first two.

Hence we have:

$$\pi(p,q) = \Pr[N[\mu,\sigma] > 0] = \Pr[N[0,1] > \frac{-\mu}{\sigma}] = 1 - \Phi[-\frac{\mu}{\sigma}]. \qquad \blacksquare$$

The data of the present model are (b, A, B). We have:

Theorem 5 'Generically,' there exists no interior voting equilibrium for the model (b,A,B).

'Proof:'

Invoking lemma 2, a voting equilibrium is a Nash equilibrium of the game with two players whose payoff functions are

$$P^A(p,q) = 1 - \Phi(-\frac{\mu}{\sigma}) - pb$$
$$P^B(p,q) = \Phi(-\frac{\mu}{\sigma}) - qb$$

where (μ,σ) is defined in the statement of the above proposition. Hence necessary first and second order conditions for an equilibrium are:

$$f(\frac{-\mu}{\sigma})\frac{d(\mu/\sigma)}{dp} = b$$
$$-f(\frac{-\mu}{\sigma})\frac{d(\mu/\sigma)}{dq} = b$$
$$-f'(-\frac{\mu}{\sigma})\left(\frac{d(\mu/\sigma)}{dp}\right)^2 + f(-\frac{\mu}{\sigma})\frac{d^2(\mu/\sigma)}{dp^2} \le 0 \quad (a),$$
$$f'(-\frac{\mu}{\sigma})\left(\frac{d(\mu/\sigma)}{dq}\right)^2 - f(-\frac{\mu}{\sigma})\frac{d^2(\mu/\sigma)}{dq^2} \le 0 \quad (b)$$

where f is the standard normal density function. The first two of these equations imply that we must have

$$\frac{d(\mu/\sigma)}{dp}+\frac{d(\mu/\sigma)}{dq}=0.$$

One may calculate that this is equivalent to the equation

$$rp-\frac{q}{r}+2p^2-3p+3q-2q^2=0,$$

where $r=\frac{A}{B}$. Viewing q as a function of p, we solve this quadratic equation for its two roots, which we denote $q^1(p;r)$ and $q^2(p;r)$. For several values of A and B, with $A>B$, I verified[6] that inequality (a) above holds if and only if $q=q^1(p;r)$ and inequality (b) holds if and only if $q=q^2(p;r)$. Furthermore, there is no value of p for which $q^1(p;r)=q^2(p;r)$. It follows, at least for these randomly chosen examples, that there is no interior equilibrium, regardless of the value of b. ■

For the example $(b,A,B)=(.1,1100,1000)$ I calculated that there is no equilibrium where either player plays at a corner. I conclude that, typically, no symmetric Kantian equilibrium exists in the voting game.

This analysis indicates that the voting paradox remains one even with perfect coordination among members of the two coalitions. Of course if b=0, then p=q=1 is a voting equilibrium – but this does not require cooperative behavior. My conclusion is that one cannot explain voting by invoking magical thinking or Kantian sentiments: rather, people must vote because they derive utility from doing so, or because they feel a

[6] This is why ‘proof’ and ‘generically’ are placed in quotation marks.

duty to vote. Here, the lesson is somewhat different from in the prisoners' dilemma. There, Kantian behavior at least rationalized limited cooperation. In the voting game, the Kantian premise gets us nowhere, because no (symmetric) Kantian equilibria exist.

4. Conclusion

Kantian equilibrium is a cooperative solution concept. It generalizes the concept of the proportional solution in production economies. It formalizes Kant's categorical imperative.

Kantian behavior is cooperative in the sense that it generates Pareto efficient allocations. The concept also demonstrates the *limits to cooperation*. We showed that it is only sometimes a Kantian equilibrium for both players in the prisoners' dilemma to cooperate for sure. And we proposed a simple model of the voting paradox, where we could not explain voting as the outcome of a non-cooperative game between the two 'sides', where each side behaves, among its members, in the Kantian fashion.

Famously, John Rawls claimed to derive is egalitarian and solidaristic political philosophy from the Kantian premise. Our formalization of that premise shows that the solidaristic behavior it implies is quite conditional – upon the rewards to cooperation being sufficiently high. Those philosophers who call themselves Kantians, and believe that this identification justifies a solidaristic world -view, I suggest, should take heed.

<u>References</u>

Elster, J. 1989. *The cement of society: A study of social order,* Cambridge University Press

Feddersen, T. J. 2004. “Rational choice theory and the paradox of not voting,” *Journal of Economic Perspectives* 18, 99-112

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

Rawls, J. 1971. *A theory of justice,* Harvard University Press

Roemer, J. and J. Silvestre, 1993. “The proportional solution for economies with both private and public ownership, *Journal of Economic Theory* 59, 426-444

Sugden, R. 1982. “On the economics of philanthropy,” *Economic Journal* 92, 341-350