June 2017 draft

"A theory of cooperation with an application to market socialism"

by

John E. Roemer
Yale University
John.roemer@yale.edu

1. <u>Man,</u> <u>the</u> <u>cooperative</u> <u>great</u> <u>ape</u>

It has become commonplace to observe that, among the five species of great ape, *homo sapiens* is by far the most cooperative.   Fascinating experiments with infant humans and chimpanzees, by Michael Tomasello and others, give credence to the claim that a cooperative protocol is wired into to the human brain, and not to the chimpanzee brain.   Tomasello's work, summarized in two recent books with similar titles (2014, 2016), grounds the explanation of humans' ability to cooperate with each other in their capacity to engage in *joint intentionality,* which is based upon a common knowledge of purpose and trust.

There are fascinating evolutionary indications of early cooperative behavior among humans.   I mention two: pointing and miming, and the sclera of the eye. Pointing and miming are pre-linguistic forms of communicating, probably having evolved due to their usefulness in cooperative pursuit of prey. If you and I were only competitors, I would have no interest in indicating the appearance of an animal that we, together, could catch and share.   Similarly, the sclera (whites of the eyes) allow you to see what I am gazing it: if we cooperate in hunting, it is useful for me that you can see the animal I have spotted, for then we can trap it together and share it.   Other great apes do not point and mime, nor do they possess sclera.

Biologists have also argued that language would likely not have evolved in a non-cooperative species (Dunbar[2009]). If we were simply cooperative, why should you believe what I would tell you? Language, if it began to appear in a non-cooperative species, would die out for lack of utility. The problem of cheap talk would be severe. In addition, language is useful for coordinating complex activities – that is, ones that require cooperation.   It would not have been worth Nature's investment in a linguistic organ, were the species not already capable of cooperation, so the argument goes.

Cooperation must be distinguished from altruism.   Altruism comes in three varieties: biological, instrumental and psychological. Biological altruism is a hard-wired tendency to sacrifice for others of one's species, which sometimes evolved through standard natural selection, as with bees and termites.   Some people speak of instrumental altruism, which is doing something to improve the welfare of another, in expectation of a reciprocation sometime in the future.   It is questionable whether this

should be called altruism at all, rather than non-myopic self-interest. Psychological altruism is caring about the welfare of others: it is a kind of preference. It is intentional, but not motivated by self-interest, as instrumental altruism is. Psychological altruism is what economists usually mean by the term.

Cooperation is not the same as psychological altruism. I may cooperate with you in building a house because doing so is the only way I can provide myself with decent shelter. It is of no particular importance to me that the house will also shelter you. Cooperation is, I believe, a more generalized tendency in humans than altruism. One typically feels altruism towards kin and close friends, but is willing to cooperate with a much wider circle. With the goal of improving human society, I think it is much safer to exploit our cooperative tendencies more fully, than our altruistic ones.

The examples I gave above of cooperation are quite primitive. Humans have, of course, engaged in much more protracted and complex examples of cooperation than hunting. We live in large cities, cheek by jowl, with a trivial amount of violence. We live in large states, encompassing millions or hundreds of millions, in peace. Early human society (in its hunter-gatherer phase) was characterized by peace in small groups, up to perhaps several hundred, but by war between groups. Our great achievement has been to extend the ambit of peaceful coexistence and cooperation to groups of hundreds of millions, between which war continues to exist. In this sense, cooperation has expanded immeasurably since early days.

Within large states, of an advanced nature, a large fraction of the economic product is pooled, via taxation, and re-allocated according to democratic decisions. We have huge firms, in which cooperation is largely decentralized. Trade unions show the extent of cooperation in firms that is decentralized and tacit when, in labor struggles, they instruct their members to 'work to rule.' In other words, it's wrong to view cooperation has primarily organized centrally; it's a false dichotomy to say that competition is decentralized and cooperation must be centrally planned. By far most instances of human cooperation are decentralized as well.

From this perspective, it is quite astonishing that economic theory has hardly anything to say about cooperation. Our two greatest contributions to understanding economic activity – the theory of competitive equilibrium, and game theory with its

concomitant concept of Nash equilibrium -- are theories of how agents compete with each other. Behavior of agents in these theories is autarchic: I decide upon the best strategy for me under the assumption that others are inert. Indeed, in Walrasian general equilibrium, a person need not even observe the *actions* that others are taking: she need only observe prices, and optimize as an individual, taking prices as given. Nothing like Tomasello's joint intentionality exists in these theories.

It would, however, be a mistake to say that economic theory has ignored cooperation. Informally, lip service is paid to the cooperative tendency of economic actors: it is commonplace to observe that contracts would not function in a cut-throat competitive society. There must be trust and convention to grease the wheels of competition. Nevertheless, this recognition is almost always in the form of the *gloss* economists put on their models, not in the *guts* of the models.

There is, however, one standard theory of cooperation, where cooperative behavior is enforced as the Nash equilibrium of a game with many stages. There are typically many Nash equilibria in such games. The 'cooperative' one is often identified as a Pareto efficient equilibrium, where the cooperative behavior is enforced by punishing players at stage $t+1$ who failed to play cooperatively at stage t. Since punishing others is costly to the punisher, those assigned to carry out punishment of deviants must themselves be threatened with punishment at stage $t+2$, should they fail to punish. Only if such games have an infinite or indefinite number of stages can this behavior constitute a Nash equilibrium. For if it is known that the game has only three stages, then no person in stage 3 will punish deviators from stage 2, because there is no stage 4 in which they would punished for shirking. So in stage 2, agents will fail to play cooperatively. By backward induction, the 'good' equilibrium unravels. (See Kandori [1992]).

What's interesting about this explanation of cooperation is that it forces cooperation into the template of *non-cooperative* Nash equilibrium. I will maintain that this is an unappealing solution, and too complex as well. It is a Ptolemaic attempt to use non-cooperative theory to explain something fundamentally different.

Let me give a simple example, the prisoners' dilemma, with two players and two strategies, *C*(ooperate) and *D*(efect). In fact, the strategy profile (*D,D*) is something

sronger than a Nash equilibrium: it's a dominant strategy equilibrium. If the game is played with an indefinite number of stages, then the behavior where both players cooperate at each stage can be sustained as a Nash equilibrium, if punishments are applied to defectors. I propose, alternatively, that in a symmetric game like this one, each player may ask himself "What's the strategy I'd like both of us to play?" This player is not considering the welfare of the other player: he is asking whether for *his own welfare* a strategy profile (*C,C*) is better than the profile (*D,D*). The answer is yes, and if both players behave according to this Kantian protocol ('take the action I'd like everyone to take'), then the Pareto efficient solution is achieved in the one-shot game.

What is needed for people to think like this? I believe it is being in a solidaristic situation. Solidarity is defined as 'a union of purpose, interests, or sympathies among the members of a group (American Heritage Dictionary).' Solidarity is not the action we take together, it is a *state of the world,* that might induce unison action. Solidarity may promote joint action, in the presence of trust: that if I take the action that I'd like all of us to take, I can trust others will take it as well. To be precise, as we will see, this behavior has good consequences when the game is symmetric (to be defined below). Symmetry is the mathematical form of 'a union of purpose or interests.' Thus Tomasello's joint intentionality, for me, is what comes about when there is a union of a solidaristic state and trust.

Trust, however, must be built up from past experience. I therefore do not claim that it is rational in a truly one-shot game to ask the Kantian question. Nash equilibrium is the rational solution of the truly one-shot game. But in real life, we are very often in situations where trust is warranted, either because of past personal experience with potential partners, or because of social conventions, of culture. In these situations, trust exists, and the Kantian question is a natural one to ask.

Now you might respond that, if the game is *really* embedded in a multi-stage game of life, then the reason that I take the action I'd like all of us to take is for fear that if, instead, I played 'Defect' (say, in the prisoners' dilemma) I will be punished in the future, or I will fail to find partners to play with me. Indeed, I think some people do think this way. But many people, I propose, do not. They have embedded the morality that playing the action I'd like all to play is 'the right thing to do' and a person should do

the right thing.   This behavior is not motivated by fear of punishment, but by morality. The morality, however, is not appropriately modeled as an object of preferences, but by a *manner of optimizing.*   This may seem like a pedantic distinction, but I will argue that it is not.

Indeed, we now come to the second way that contemporary economics explains cooperation, and that is under the rubric of *behavioral economics.*   Behavioral economics has many facets: here I am only concerned with its approach to explaining cooperation. I claim that the general strategy adopted by behavioral economists to explain cooperation is to insert *exotic arguments* into preferences – like a sense of fairness, a desire for equality, a care for the welfare of others – and then to derive the 'cooperative' solution as a *Nash* equilibrium of this new game.   Thus, for example, a player in the prisoners' dilemma plays *C* because it would be unfair to take advantage of an opponent playing *C* by playing *D*.   In this formulation both (*C,C*) and (*D,D*) would be Nash equilibria, if I incur a psychic cost for playing *D* against your *C*.   Or suppose we simply say that the player gets a 'warm glow' from playing *C* (see Andreoni [1990]). Then the unique Nash equilibrium, if the warm glow is sufficiently large, will be (*C,C*).

Indeed, Andreoni's 'warm glow' merits further comment. I think it's true that many people get a warm glow from playing the Kantian action, from doing the right thing. But the warm glow is an unintended side effect, to use Elster's (1981) terminology, not the reason for the action.   I teach my daughter the quadratic formula. She gets it: I get a warm glow. But I did not teach her the formula *in order* to generate the warm glow, which came along as a result that I did not intend.   Andreoni has reversed cause and effect.   The same criticism applies to explanations of charity.   The Kantian explanation is that I give what I'd like everyone in my situation to give, rather than my giving because it makes me feel good – which is not to deny that I do feel good when I do the right thing.

The Kandori explanation of cooperation as a Nash equilibrium in a multi-stage game with punishments is what Elster (1989) calls a *social norm.*   To be precise, it is part of Elster's characterization of social norms that those who deviate be punished by others, and those who fail to punish deviators are themselves punished by others. Doubtless, many examples of cooperation are social norms: but not all are.   It has often

been observed by economists that normal preferences for risk will not explain the extent of tax compliance, given the probabilities of being caught for evading, and the subsequent (small) fines. In some countries, tax evaders' names are published in the newspaper, and there it may well be that compliance is a social norm. In many cities, large numbers of people recycle their garbage. Often, nobody observes whether or not one recycles. There is no punishment, in these cases, for failing to recycle: but many recycle nevertheless. Assuming that recycling is somewhat costly, the Nash equilibrium – even if people value a clean environment – is not to recycle. (I should not recycle if the cost of recycling to me is greater than the marginal contribution my recycling makes to a clean environment.) Recycling, I think, is better explained as a Kantian equilibrium. Not everyone recycles because not everyone thinks Kantian.

People's trust in others may come with thresholds. I will recycle if I see or read that fraction q of my community recycles. There is a distribution function of the thresholds q in the community. In figure 1 such a distribution function is graphed; there is a stable equilibrium where fraction q^* recycle. (There are also unstable equilibria where fraction 0 or fraction 1 recycle.) I have called these people conditional Kantians; Elster (in press) calls them quasi-moral (if $1 > q > 0$) and reserves the label 'Kantian' for those for whom $q = 0$. Nash players have $q = 1$: they always play Nash, no matter how many others are playing Kantian.

[place figure 1 here]

There are, I think, three explanations of how workers cooperate when they go on strike, or why people join revolutionary movements or dangerous demonstrations. The first, promulgated by Olson (1965), is of the repeated-game-with-punishments variety. Workers who cross the picket line are beaten up. Or, there is a carrot: joining the union comes with side payments. Olson's explanation is clearly cooperation-as-a-Nash equilibrium. Recently, Barbera and Jackson (2017) explain these actions as occurring because participants enjoy an expressive value from the action: they value expressing their opposition to the regime or the boss. This is what I've called the behavioral-economics approach: putting exotic arguments into preferences. I (in press) model

strikes as games where players' strategies are their probabilities of striking: in the case where all preferences are the same, the Kantian equilibrium is a probability that will maximize my expected income if everyone strikes with that probability. (With heterogeneous preferences, the story is more complicated.) Preferences are straight-forward economic preferences, with no expressive element; nor are there punishments. It is not unusual in this model for the Kantian equilibrium to be that each strikes with probability one.   In reality, we do not often observe this, because not everyone is a Kantian.   There may well be punishments, in reality, to deter those who would not strike when the strike is on.   But it is wrong to infer that those punishments are the reason that *most* people strike. They may be needed only to control a fairly small number of Nash optimizers.   And if the workers are conditional Kantians, or *q*-Kantians, it is possible for a strike to unravel if even a small number of Nash players are not deterred from crossing the picket line.

I have thus far only discussed symmetric games – true solidaristic situations, where all payoff functions are the same, up to a permutation of the strategies. I will now formalize what I have proposed, before going on to the more complicated problem of games that are not symmetrical, where payoff functions are heterogeneous.

12. Simple Kantian equilibrium

Consider a game where all players choose strategies from an interval *I* of real numbers, and the payoff function of player *i* is

$$V(E^i, \mathbf{E}^{-i}) \tag{2.1}$$

where the vector of strategies is denoted $\mathbf{E} = (E^1,...,E^n)$ and $\mathbf{E}^{-i}$ is the vector $\mathbf{E}$ without its *i*th component.   This is a symmetric game.

<u>Definition</u> 2.1   A *simple Kantian equilibrium* is a vector $\mathbf{E}^* = (E^*, E^*,...,E^*)$ such that:

$$E^* = \arg\max_{E \in I} V(E; E,...,E) \ . \tag{2.2}$$

That is, among all vectors where everyone plays the same strategy, the strategy that everyone would choose is E^* .   This is the formalization of the statement that everyone plays the strategy he'd like everyone to play.

<u>Definition</u> 2.2   A symmetric game is *strictly monotone increasing* if each player's payoff is strictly monotone increasing in the strategies of the other players. It is *strictly monotone decreasing* if it is strictly monotone decreasing in the strategies of the other players.

The standard example of a strictly monotone increasing game is when a person's *E* is contribution to a public good. The more others contribute, the higher my welfare. The standard example of monotone decreasing game is the tragedy of the commons. We all fish on a common lake, and the more others fish, the less productive is the lake for me.

The Nash equilibria of strictly monotone games are Pareto inefficient.   In the case of monotone increasing games, in Nash equilibrium, people contribute too little; in the case of monotone decreasing games, they fish too much. But we have:

<u>Proposition</u> 2.1   *The simple Kantian equilibrium of a strictly monotone game is Pareto efficient.*

This result is what I referred to earlier when I said that in a symmetric game, if everyone plays the strategy he'd like everyone to play, the result is 'good.'   Because this result is so central to the idea of Kantian optimization, I will prove it here.

<u>Proof</u> <u>of</u> <u>Proposition</u> 2.1:
Let the game be strictly monotone decreasing. Let $(E^*,...,E^*)$ be a simple Kantian equilibrium. If it is not Pareto efficient, then there is a vector $\mathbf{E} = (E^1,...,E^n)$ such that:
$$(\forall i)(V(E^i;\mathbf{E}^{-i}) \geq V(E^*;E^*,...,E^*)) \qquad (2.3)$$
where the inequality is strict for at least one index *i*.   Let *j* be an index such that $E^j = \min_i E^i$.   Then $V(E^j;E^j,....,E^j) > V(E^j;\mathbf{E}^{-j})$.   This is because we have reduced the efforts (*E*) of some players other than *j* and increased the efforts of no players, because E^j is minimal among the $\{E^i\}$. So the strict inequality just stated follows from the fact that the game is strictly monotone decreasing. But this last inequality implies that:
$$V(E^j;E^j,...,E^j) > V(E^*;E^*,...,E^*) \qquad (2.4)$$

contradicting the assumption that **E*** is a simple Kantian equilibrium. This contradiction proves the proposition.

An analogous proof establishes the claim for monotone increasing games. ∎

In this sense, simple Kantian equilibrium resolves the inefficiencies due to both negative and positive externalities that are characteristic of Nash equilibrium.

The general version of a 2×2 symmetric prisoners' dilemma is given by the payoff matrix in table 1:

	C	D
C	(0,0)	(-c,1)
D	(1,-c)	(-b,-b)

Table 1.   The payoffs are (row player, column player).

where
$$0 < b < c . \tag{2.5}$$
Let's look at the more complicated version of the PD where each player plays a mixed strategy: with probability p the row player plays C, and with probability q the column player plays C.   Then the payoff function of the row player is:
$$V^{row}(p,q) = -b(1-p)(1-q) - cp(1-q) + (1-p)q , \tag{2.6}$$
and by symmetry, the payoff function of the column player is:
$$V^{col}(p,q) = -b(1-p)(1-q) + p(1-q) - c(1-p)q . \tag{2.7}$$
Calculate that :

$$\frac{dV^{row}}{dq} = (b+1)(1-p) + pc > 0 \ \text{ for all } p, \tag{2.8}$$

so this is a strictly monotone increasing game.   Hence the simple Kantian equilibrium of the game is Pareto efficient, by Proposition 2.1.   This means there is no mixed-strategy pair that can give a higher expected utility to both players than the simple Kantian equilibrium.   The exact form of the simple Kantian equilibrium of the game depends

upon the values of b and c. The Nash equilibrium of the game is always $(0,0)$: both players defect for sure, and the outcome is inefficient.

Let me now make precise one of my criticisms of the behavioral economics. One can compute that if $c < 1$, then the simple Kantian equilibrium of the PD game is that both players cooperate with probability:

$$p^* = \frac{2b+1-c}{2(1+b-c)} \, , \tag{2.9}$$

a probability strictly between zero and one. In other words, *full cooperation* ($p = 1$) is *not* the simple Kantian equilibrium! Of course, it remains true that (p^*, p^*) is Pareto efficient: in particular, it is better for both players (in ex ante utility) than the strategy profile $(1,1)$. Now suppose you are a behavioral economist, and you want to insert an exotic argument into the preferences of the agents so that the *Nash* equilibrium of the game will be (p^*, p^*) ? How would you do it? Recall that the preferences you create must not work just for *this* PD game but for *all* PD games.

There turns out to be one, and I believe only one, sensible way of doing this. Assign each player a new payoff function, which is the *sum* of the payoff functions of the two players in the original PD. That is, define:

$$Q(p,q) = V^{row}(p,q) + V^{col}(p,q) \, . \tag{2.10}$$

Now consider the game where both players have the payoff function Q – of course, the row player continues to play p and the column player q. It is not hard to check that the *Nash* equilibrium of this game is indeed:

$$\hat{p} = \hat{q} = p^* = \frac{2b+1-c}{2(1+b-c)} \, . \tag{2.11}$$

So we *can* rationalize the simple Kantian equilibrium of the PD game as the Nash equilibrium of a game where each player is maximizing an 'exoctic' preference order, the *total* payoff.

This might appear to support the behavioral economist's strategy. Here indeed – and this can be checked – the simple Kantian equilibrium of a symmetric game is always

a Nash equilibrium of an altered game where each player's payoff function is the *sum* of all players' payoff functions in the original game.   There are, however, two problems with this move: first, is it credible to think that's what players are doing when they play the cooperative strategy in such a game – that they are attempting to maximize the total payoff?   This is something that can be studied experimentally, and I conjecture it will not be borne out.   But secondly, this trick (of transforming a Kantian equilibrium into a Nash equilibrium of a game with altered preferences) only works when the game is symmetric. A bit more on this later.

3. Simple production economies

We now consider a class of simple production economies in which we can study Kantian optimization in environments with negative and positive externalities. Suppose there is a production function that produces a desirable consumption good from the efforts of individuals, according to the production function $G(E^1,...,E^n)$, where G is concave, increasing and differentiable. E^i is the labor expended by person i, measured in efficiency units.   We denote $E^S = \sum E^i$.   Player i has a utility function $u^i(x,E)$ where x is the consumption good produced.   As usual u is concave, increasing in x , decreasing in E, and differentiable.   For the moment, we allow the preferences of individuals to differ.

A. The fishing economy

In the fishing economy, we think of G as production of fish from a lake, which suffers from congestion externalities, the more labor is expended in fishing: hence, the concavity of G.   Recall that an interior allocation is Pareto efficient exactly when :

$$\text{for all } i, \quad -\frac{u_2^i(x^i,E^i)}{u_1^i(x^i,E^i)} = G'(E^S) , \tag{3.1}$$

for this is the statement that the marginal rate of substitution between labor and fish for each fisher (MRS^i) equals the marginal rate of transformation of labor into fish (*MRT*).

The *allocation rule* in this economy is given by:

$$x^i = \frac{E^i}{E^S} G(E^S) \; ; \tag{3.2}$$

that is, except for random noise, each fisher receives fish in proportion to the efficiency units of labor he expends. Another way of putting this is that 'each fisher keeps her catch.'

We can define a game: the payoff to fisher i, is given by the function:

$$V^i(E^1,...,E^n) = u^i(\frac{E^i}{E^S} G(E^S), E^i) \; . \tag{3.3}$$

It is well-known, that due to the strict concavity of G, the *Nash* equilibrium of this game is Pareto inefficient. This is perhaps the classical example of the tragedy of the commons. In Nash optimization, players do not take into account the negative externality they impose on other fishers by their own fishing. For each addition hour of I fish, I lower the marginal productivity of everyone's labor on the lake. Indeed, the Nash equilibrium is the solution of these equations:

$$\text{for all } i, \quad -\frac{u_2^i}{u_1^i} = \frac{E^i}{E^S} G'(E^S) + (1 - \frac{E^i}{E^S})\frac{G(E^S)}{E^S} \; . \tag{3.4}$$

In words, the marginal rate of substitution of a fisher is equal to a convex combination of the marginal rate of transformation and the average product. Only in the case of one fisher is this allocation Pareto efficient.

To apply the concept of simple Kantian equilibrium, we will assume for the moment that $u^i = u$ for all i. What is the simple Kantian equilibrium of the game? It is given by solving:

$$\frac{d}{dE} u(\frac{1}{n} G(nE), E) = 0 \; , \tag{3.5}$$

for the solution of (3.5) is the amount of fishing time that each player would like all to expend. Taking the derivative, we have:

$$\frac{d}{dE} u(\frac{1}{n} G(nE), E) = u_1 \frac{1}{n} G'(nE)n + u_2 = 0$$

$$\text{or} \quad G'(E^S) = -\frac{u_2}{u_1}. \tag{3.6}$$

But this is the condition for Pareto efficiency! Hence the simple Kantian equilibrium is Pareto efficient.

Now, a caveat. You might have observed that the game defined in (3.3) is a monotone decreasing game, and hence concluded from Prop. 2.1 that the simple Kantian equilibrium (SKE) is Pareto efficient, and so the derivation (3.6) is redundant. But that inference is false. What Proposition 2.1 shows is that the SKE is Pareto efficient *among all allocations that can be achieved in the game, so defined :* that means among all allocations in which consumption of fish is proportional to labor expended. But (3.6) is a much stronger statement: it says that the SKE is Pareto efficient in the set of *all feasible allocations* for this economy.   There is no way of allocating the total fish caught to the fishers, through any intricate system of redistribution, that can Pareto dominate the SKE of the fishing economy.   In other words, we achieve full efficiency even though we restrict ourselves to allocations where each fisher keeps her catch. The demonstration in (3.6) is stronger than Proposition 2.1.

Again, I must make the comparison with the behavioral-economics approach. Because this is a symmetric game (when all utility functions are the same), it is indeed the case that the *Nash* equilibrium of the game where each player maximizes the sum of utilities of all players is the simple Kantian equilibrium given by (3.6).   So we have, at this point, two explanations if a fishing community achieves the Pareto efficient allocation in which each keeps his catch: either each is fishing the amount of time he would all everyone to fish, or each is a complete altruist, desiring to maximize total utility.

B. Heterogeneous preferences

We now relax the assumption that all utility functions are the same, and assume the profile of utility functions is $(u^1,...,u^n)$.   One simple way to generate heterogeneous preferences is to say that everyone has the same preferences over fish and *labor time*, but since fishers have different skill levels, this induces heterogeneous preferences over fish and efficiency units of labor.   Recall that the relevant labor in our models is the latter.

It's now the case that simple Kantian equilibria will generally not exist: the labor that I would most like everyone to expend is the different from the labor you would most like everyone to expend.   There is, however, a generalization of Kantian optimization that we can use with heterogeneous preferences.

Suppose at an allocation $\mathbf{E} = (E^1,...,E^n)$ I am contemplating increasing my fishing time by 5%. I ask myself: How would I like it if everyone increased her fishing time by 5%? The implicit moral commandment here is that I should only increase my fishing time by 5% if I'd be happy were everyone to do the same. It's a way of making me take into account the negative externality that is imposed by my contemplated action. But I don't ask myself how would my increased fishing affect others, rather *how, if they emulated my action, their increased fishing times would affect me.* Don't worry too much at this point about the psychological realism of this question. Instead, let's study the properties of an equilibrium when everyone thinks in this way.

Definition 3.1 An allocation of efforts $\mathbf{E} = (E^1,...,E^n)$ is a *multiplicative Kantian equilibrium* (a $K^\times$ equilibrium) if *nobody* would like to rescale *everybody's* fishing time by *any* non-negative scale factor. That is:

$$(\forall i)\left(\arg\max_{r \geq 0} u^i(\frac{rE^i}{rE^S}G(rE^S), rE^i) = 1 \right) \qquad (3.7)$$

I have stated the definition for the fishing economy, but in the attached footnote, I state it for a general game in normal form.[1]

We have the general result, using the definition in the footnote:

Proposition 3.1 *If* $\mathbf{E} = (E^1,...,E^n)$ *is a* $K^\times$ *equilibrium of a strictly monotone game* $\{V^i\}$ *, then it is Pareto efficient in the game.*

The proof of this proposition is very similar to the proof of proposition 2.1.

Proposition 3.1 justifies calling the $K^\times$ equilibrium a model of cooperation, because it (also) solves the two major kinds of inefficiency afflicting Nash equilibrium (competitive behavior): Pareto inefficiency in environments with positive externalities,

---

[1] Let the payoff functions of the game be $\{V^i\}$ where the strategy space for each player is an interval of real numbers. Then $(E^1,...,E^n)$ is a $K^\times$ equilibrium of the game when $(\forall i)\left(\arg\max_r V^i(rE^1,...,rE^n) = 1 \right)$.

as in the provision of public goods, and inefficiency in the presence of negative externalities, as in tragedies of the commons.

In particular, the fishing game of (3.7) is a strictly monotone decreasing game. So Proposition 3.1 implies that the multiplicative Kantian equilibrium, if it exists, is Pareto efficient in the game. As before, we must recall that being efficient 'in the game' means in the class of allocations where each keeps his catch. But we could ask for a stronger result: is the $K^\times$ equilibrium efficient in the *economy*?

We have:

Proposition <u>3.2</u> *Any positive* $K^\times$ *equilibrium of the fishing game is Pareto efficient in the economy.*

Again, because this is a key result, let's prove it.

<u>Proof:</u>

1. By concavity of the utility functions and G, it is only necessary to show that :

$$(\forall i)\left(\frac{d}{dr}\Big|_{r=1} u^i(\frac{E^i}{E^S}G(rE^S),rE^i) = 0\right) \Rightarrow \text{ the allocation is Pareto efficient.}$$

This suffices, because the antecedent to the implication sign is the first-order characterization of $K^\times$ equilibrium.

2. Compute that:

$$\frac{d}{dr}\Big|_{r=1} u^i(\frac{E^i}{E^S}G(rE^S),rE^i) = u_1^i \cdot \frac{E^i}{E^S}G'(E^S)E^S + u_2^i E^i = 0 \ . \tag{3.8}$$

Using the assumption that $E^i > 0$, this equation reduces to:

$$(\forall i)\left(-\frac{u_2^i}{u_1^i} = G'(E^S)\right) , \tag{3.9}$$

which is exactly the condition for Pareto efficiency at an interior solution. ∎

As with the case of homogeneous preferences, we can therefore assert a stronger statement than Proposition 3.1: in the fishing economy, multiplicative Kantian

optimization resolves the tragedy of the commons that afflicts Nash equilibrium. (Recall the Nash equilibrium of the fishing game is given by (3.4).)

Do such equilibria exist? These are allocations in which fish consumed is proportional to labor expended *and* the allocation is Pareto efficient. The idea of looking for such allocations is due to Joaquim Silvestre. In Roemer and Silvestre (1993), we proved that in economies much more general than the simple production economies studied here, such allocations always exist. Silvestre's question was asked in the context of thinking about the socialist ideal: an allocation in which the total product is distributed in proportion to labor expended, *and* is Pareto efficient. Remarkably, perhaps, this question had not been raised earlier by mathematical socialist economists like Oscar Lange and Michio Morishima. It was, however, until several years later that I observed that these allocations, which Silvestre and I called *proportional solutions*, had the property of being stable with respect to the kind of optimization that I now call Kantian (see Roemer [1996, Theorem 6.6]). Also, as an illustration of the circuitousness of one's thinking, the much simpler idea of simple Kantian equilibrium in symmetric games did not occur to me for another 20 years.

We can now repeat our earlier question: Is there a way of altering preferences in the fishing game so that the Nash equilibrium of the *altered* game is the Pareto efficient allocation in which each keeps his catch (i.e., the multiplicative Kantian equilibrium)? (There is generally a unique such allocation; in any case, there is a finite number of such allocations.) The answer is there is no simple way of doing this: the idea of giving all players the desire to maximize total utility no longer works. For further discussion, I refer the reader to Roemer (in press, chapter 6).

In other words, in games that are fairly complicated, the behavioral economist's strategy of altering preferences in order to achieve the good (cooperative) solution as the *Nash* equilibrium of a game does not work. Essentially, one has to know what the good solution is *a priori* and then one can jimmy preferences to give the right answer under Nash optimization. But this cannot be regarded as decentralizing cooperation. In contrast, multiplicative Kantian optimization decentralizes cooperation, in the exact sense that Nash optimization decentralizes competition. As with Nash, there is no obvious

solution to the dynamic problem of how one converges to a Nash (or Kantian) equilibrium. But, as with Nash, both equilibria are stable with respect to kind of optimization that players are using, once it is achieved.

I will have more to say about the realism of proposing that players in a game might optimize in the multiplicative Kantian fashion – but later. At this point, however, I still want to explicate the properties of this kind of thinking.

C. The hunting economy

Consider the game that according to anthropologists characterized early hunting societies. Unlike fishers, hunters divided the catch equally (more or less). A group of hunters go out into the bush for a few days. They return with their catch, and divide it equally among the group.

Here the allocation rule is:

$$x^i = \frac{G(E^1,...,E^n)}{n} , \quad (3.10)$$

which is the *equal-division* rule. The game defined by the equal-division rule is given by:

$$V^{ED\,i}(E^1,...,E^n) = u^i(\frac{G(E^S)}{n},E^i) . \quad (3.11)$$

This is a game with positive externalities: the total catch is a public good. The game is strictly monotone *increasing*. Again, the Nash equilibrium, which is given by the next set of equations, is Pareto inefficient as long as there is more than one hunter:

$$-\frac{u_2^i}{u_1^i} = \frac{G'(E^S)}{n} . \quad (3.12)$$

Why is the Nash equilibrium inefficient? Because a hunter might like to take a nap for several hours under a bush, given what others are doing. He contributes less to the public good than efficiency requires.

It turns out that multiplicative Kantian optimization does not resolve this inefficiency. But there is a kind of Kantian optimization that does. Suppose the hunter asks himself, "I'd like to take a two hour nap. But how would I feel if everyone took at two hour nap?" If the moral commandment is to take the nap only if the answer to this questio is affirmative, then an equilibrium is defined as follows:

Definition 3.2 An allocation $\mathbf{E} = (E^1, ..., E^n)$ is an *additive Kantian equilibrium* (K^+) if and only if:

$$(\forall i)\left(\arg\max_r u^i(\frac{G(E^S + nr)}{n}, E^i + r) = 0 \right). \tag{3.13}$$

In other words, *nobody* would like to *translate* the effort vector by *any* number (positive or negative)[2].

I will now leave it to the reader to verify:

Proposition 3.3 *Any additive Kantian equilibrium of the fishing game is Pareto efficient in the economy.*

Thus, additive Kantian optimization resolves the under-provision of the public good in the hunting economy.

Kantian optimization always takes the form of each asking: "If we all take a similar symmetric action, what would I like that action to be?" In the case when the game is symmetric, the action is 'making the same contribution.' In the fishing game, the action is 'rescaling the current allocation by a common factor;' in the hunting game, it is 'translating the current allocation by a common factor.'

D.   More general allocation rules in simple production economies

Let's define an allocation rule for an economic environment specified by the data $(u^1, ..., u^n, G)$ as the share of output θ^i each individual receives as a function of the vector of effort contributions. We have studied two rules, the proportional and equal-division rules, whose share rules are:

---

[2] For general games, $(E^1, ..., E^n)$ is an additive Kantian equilibrium if:

$$(\forall i)\left(0 = \arg\max_r V^i(E^1 + r, ..., E^n + r) \right).$$

$$\theta^{\text{Pr.}}(E^1,...,E^n) = \frac{E^i}{E^S}, \qquad \theta^{ED,i}(E^1,...,E^n) = \frac{1}{n} \text{ for all } i \ . \qquad (3.14)$$

Indeed, these are surely the two classical notions of fair division of a jointly produced output. We have shown that each of these rules can be efficiently implemented -- more strongly, efficiently decentralized—by a specific kind of Kantian optimization.

We can also think more generally about kinds of Kantian symmetrical treatment. Define a *Kantian variation* as a function $\varphi(E,r)$ where E is a contribution and r is a number where the following holds:

$$\varphi(E,1) \equiv 1, \quad \varphi(E,r) \text{ is increasing in } r \ . \qquad (3.15)$$

The *multiplicative Kantian variation* is given by:

$$\varphi^{\times}(E,r) = rE \qquad (3.16)$$

and the *additive Kantian variation* is given by:

$$\varphi^{+}(E,r) = E + r - 1 \ . \qquad (3.17)$$

Now we say that the allocation rule θ is *efficiently implemented by the Kantian variation* φ on the set of economies $\{(u^1,...,u^n,G)\}$ if for all economies:

$$(\forall i)\left(1 = \arg\max_r u^i(\theta^i(\varphi(\mathbf{E},r)G(\sum \varphi(E^i,r))),\varphi(E^i,r))\right) , \qquad (3.18)$$

where $\varphi(\mathbf{E},r) \equiv (\varphi(E^1,r),...,\varphi(E^n,r))$.

Equations (3.18) simply state the generalization of multiplicative and additive Kantian equilibrium to other allocation rules, and other types of Kantian variation. They say that an allocation has the property that, for the given allocation rule θ, nobody would like to vary the entire effort vector according to any transformation permitted by the variation defined by φ.

Definition 3.2   The pair (θ,φ) is an *efficient Kantian pair* if the allocation rule θ is efficiently implemented as a K^{φ} equilibrium on all economies of the form $(u^1,...,u^n,G)$.

What we have thus far shown is that $(\theta^{\text{Pr}}, \varphi^{\times})$ and $(\theta^{ED}, \varphi^{+})$ are efficient Kantian pairs. The question is: Are there any others?

The answer is that there is a unidimensional continuum of such pairs, of which the two we have studied are the endpoints. Consider allocation rule of the form:

$$\theta^{\beta,i}(E^1,...,E^n) = \frac{E^i + \beta}{E^S + n\beta} \tag{3.19}$$

where $0 \leq \beta < \infty$. Notice that θ^0 is the proportional rule, and that as β approaches infinity, θ^β approaches the equal-division rule. Because of this, let's define $\theta^\infty = \theta^{ED} = \frac{1}{n}$. We have:

Proposition 3.4 *For every rule θ^β , there is a Kantian variation that implements it efficiently on the domain of production economies we are considering. Furthermore, there are no other allocation rules that can be efficiently implemented on this domain with respect to any Kantian variation[3].*
(For proof, see Roemer (in press, proposition xx).)

What do the rules θ^β look like? It is easy to show that they are 'convex combinations' of the proportional rule and equal division rule. Let's solve the following equation for λ :

$$\frac{E^i + \beta}{E^S + n\beta} = \lambda \frac{E^i}{E^S} + (1-\lambda)\frac{1}{n}. \tag{3.20}$$

The solution is $\lambda = \frac{E^S}{E^S + n\beta}$. The important fact is that the value of λ is independent of i. This means that the rule θ^β is simply this: it takes a share λ of the output $G(E^S)$ and distributes it to the participants in proportion to the efficiency units of labor

---

[3] There are two caveats. If $\beta < 0$, the allocation rules can also be efficiently implemented; however, Kantian allocations may not exist. And for $\beta = 0$ we must insist that the allocation be strictly positive. (The zero vector is a multiplicative Kantian equilibrium but it is not Pareto efficient.)

expended, and it distributes the rest of the output equally to all. The share λ , however, *depends upon the equilibrium*. We know that as β travels from zero to infinity, the share λ travels from one to zero: but that is all that we can say. It is not even clear that the share that is distributed in proportion to effort is monotone decreasing in β. (I have not studied this question, but I would suppose that it is not generally so.)

It's for this reason that I put 'convex combination' in quotations marks earlier. We cannot specify the share λ to be – say – one-half, and then quickly choose the right allocation rule to implement that share. To say this mathematically, the mapping from λ to β is complicated, as to know it we have to compute the equilibrium to know E^S ! Only after knowing the equilibrium for each β can we find the one that implements a particular combination of proportional and equal division.

The upshot of this discussion is that the class of allocation rules that can be efficiently implemented by *any* kind of Kantian thinking is precisely the class of rules 'generated' by the two classical egalitarian methods of distribution: in proportion to effort and equally. As these rules have a venerable history as rules of fair allocation, Proposition 3.4 re-enforces the view that Kantian optimization is the right way of thinking about morality and, may we say , cooperation.

I conclude this section with a remark about the behavioral feasibility of Kantian optimization. I think simple Kantian optimization is natural – there are many symmetric games that describe our social interactions, and in those games, I think many people ask the Kantian question: what's the action I'd like everyone to take? Multiplicative and additive Kantian optimization are, however, not natural. I think symmetry plays a large role in our conceptions of fairness, and the symmetrical treatment of all characteristic of Kantian optimization recommends it as a way of implementing fairness. But I do not claim societies have discovered these complex ways of optimizing. Rather, I see the approach as prescriptive. If the fairness involved in these kinds of Kantian optimization appeals to people as implementing fairness, we may recommend –- to a fishing community, for example – the multiplicative Kantian equilibrium as a solution of their problem. Doing so, of course, would require the planner to know the preferences of the fishers, unless some process I do not yet understand could lead dynamically to the Kantian equilibrium of the game.

4. Market socialism

Thus far, we have studied economies that do not trade on markets. Indeed, there is very little that can be thought of as comprising trade in the fishing and hunting economies. We now ask what role Kantian optimization can play in market economies.

Indeed, I will propose that Kantian optimization can play an important role in a design for market socialism. The 'design problem' for socialism is stated by the philosopher G.A. Cohen as follows:

> In my view, the principal problem that faces the socialist ideal is that we do not know how to design the machinery that would make it run. Our problem is not, primarily, human selfishness, but our lack of a suitable organizational technology: our problem is a problem of design. It may be an insoluble design problem, and it is a design problem that is undoubtedly exacerbated by our selfish propensities, but a design problem, so I think, is what we've got. (Cohen, 2009)

What Cohen means is this. Capitalism has a design consisting of private property rights in labor and productive assets, and free trade, mediated by prices that equilibrate supply and demand. The *ethos* that makes capitalism work is the maximization of self-regarding preferences in an autarchic manner: the individualist ethos is captured not only in the nature of preferences, but in the autarchic manner in which people optimize (à la Nash). Cohen also views the motivation for economic activity under capitalism as being 'greed and fear,' a point with which I do not completely agree.

Socialism, however, is supposed to be an economic system characterized by cooperation. The natural question becomes, how can one design an economic system based on cooperation to deliver good results? The first theorem of welfare economics for capitalist economies is the main example of capitalism's good result. Can we design a mechanism, that given a socialist ethos, would deliver good results? A good result for socialism might demand a higher standard than for capitalism: we would desire to have not only efficiency but a substantial degree of *income equality*.

Cohen, in the above quotation, says that it is not primarily (selfish) human nature that poses a problem for socialism, it is the lack of a solution to this design problem.   Cohen, evidently, is optimistic about human nature – and there is indeed good reason to be so. Many people, in today's capitalist societies, do not design their lives to maximize their wealth, but to do work that is useful for society. I agree with Cohen that human nature is not the primary obstacle to socialism, but the lack of a design that can *decentralize economic activity* to achieve good results, *given* that a socialist ethos exists in the population.

To state the problem slightly differently, recall the huge effect that Marx's theory of historical materialism had on engendering an international socialist movement. Marx offered – if not a design in the sense here being discussed – a *vision* of the feasibility and indeed historical necessity of socialism.   We need not here debate the validity of that vision: what's salient is that, possessing this vision, millions of people organized to attempt to realize it.   Now that vision has soured, due to the experiences of twentieth century socialism, polluted and hindered as they were by political dictatorship and the fear of introducing markets.   Having a *design* for how socialism could work, *given* a willing population, is of primary importance in rekindling the socialist vision.

I am less suspicious of markets than Cohen was, and so I will propose how incorporating *cooperative optimization* into a market economy can produce eqiuilibria which are *decentralized, efficient, and equitable.*

The model is considerably more general than the one I now exposit, for purposes of clarity.   Assume there is one produced good, which is used both for investment and consumption.   The good is produced from capital and labor, according to a production function $G(K,L)$, which is operated by a firm.   The firm is owned by private citizens and the state.   The share in the firm's profits of private citizen i is θ^i , and the state's share is θ^0 : so $\sum_{i=0}^{n} \theta^i = 1$.   Individuals, as well as owning shares of the firm, have endowments of efficiency units of labor in the amount ω^i .   Citizen i has a preference order over consumption of the good and labor expended, denoted in efficiency units, represented by a utility function $u^i(x,E)$.   The state is endowed with an amount of

capital K_0. Individuals own no capital, nor do they hold inventories of the consumption good. (The capital good and consumption good are, as I said, the same good.)

So far, I have described exactly the set-up of an Arrow-Debreu economy, except for the state's partial ownership of the firm. There will be three prices in this economy, a price for output p, an interest rate for capital r, and a wage w per unit of efficiency labor. The firm behaves exactly as an Arrow-Debreu firm: it demands capital and labor and supplies the good in order to maximize profits. All incomes in the economy, except the state's revenues, will be taxed a rate $t \in [0,1]$, and the revenues will returned to the citizenry as a demogrant: that is, divided equally among them. This is a flat, or affine, income tax.

How does the state behave? It supplies capital to the firm to maximize its income, which it uses to purchase the good produced by firm, to be used for next period's investment. In fact, the model is truncated: there is only one period, for purposes of simplicity. So the firm demands capital, which is uniquely supplied by the state, and supplies the state with the good, which the state purchases with its revenues, consisting of its interest on capital loaned and its share of the firm's profits.

How do consumer-workers behave? Given a worker's labor supply E^i to the firm, it uses its revenues to purchase the good for consumption purposes. Workers' revenues come from two sources: their after-tax wage and profit income, and the demogrant.

The only substantial way in which the model differs from the standard Arrow-Debreu model is in the determination of the labor supplies of workers. The vector of labor supplies $(E^1,...,E^n)$ must be an *additive Kantian equilibrium* of a game to be defined below.

Finally, a price vector (p,w,r) comprises a *Walras-Kant equilibrium* if, given the optimizing behaviors described, all markets clear: that is, the supply of the good by the firm equals the demand of consumers for the good plus the demand by the state for the good (investment), the supply of capital by the state equals the firm's demand for capital, and the total efficiency units of labor supplied by workers equals the demand for labor by the firm.

We must now be more precise in order to define the game of which the effort vector must be an additive Kantian equilibrium. First, define the income of the state, which will be:

$$I^0(p,r,w,E^S) = \theta^0 \Pi(K_0,E^S) + rK_0 \ , \tag{4.1}$$

where the firms profits are:

$$\Pi(K_0,E^S) = pG(K_0,E^S) - wE^S - rK_0 \ . \tag{4.2}$$

In words, (4.1) states that the state's income consists of its share of firm profits plus the interest on its investment, which will in equilibrium be its endowment of capital.

The income of the worker is given by:

$$I^i(E^1,...,E^n) = (1-t)wE^i + (1-t)\theta^i\Pi(K_0,E^S) + \frac{t}{n}(pG(K_0,E^S) - I^0(K_0,E^S)) \ . \tag{4.3}$$

The first term on the right-hand side of (4.3) is agent i's after-tax wage income, the second term is her after-tax profit income, and the third term is the demogrant, which is the agent's per capita share of tax revenues, where taxes are levied on all incomes except the state's.

The payoff function for worker i is given by :

$$V^i(E^1,E^2,...,E^n) = u^i\left(\frac{I^i(E^1,...,E^n)}{p},E^i\right) \ . \tag{4.4}$$

That is, the worker's utility is generated by spending her entire income on the consumption good, and her labor.

The full definition of a Walras-Kant equilibrium is:

Definiton 4.1 A *Walras-Kant equilibrium at a tax rate t* is

- a price vector (p,w,r) ,
- demands for capital and labor by the firm in amounts (K,D) and an output of the good $G(K,D)$ that maximizes the firm's profits at (p,w,r);

- demands for the good $x^0, x^1, ..., x^n$ by the state and the n consumers, such that $px_0 = I^0(K_0, E^S)$ and , for all $i \geq 1$, $px^i = I^i(E^1, ..., E^n)$;

- supplies of labor by workers to the firm in amounts $(E^1, ..., E^n)$ that comprise an additive Kantian equilibrium of the game among workers defined by the payoff functions (4.4);

- a supply of capital to the firm by the state in amount K_0 (which necessarily maximizes the state's income, given prices)

such that:

- all markets clear; that is:

  - $K_0 = K$

  - $E^S = D$

  - $G(K, D) = \sum_{i=0}^{n} x^i$.

(The third bullet point says that each agent expends its entire income on the good.) If we assume that the depreciation rate of capital is zero, then the state's capital, to be used in the next period, will be $K_0 + \dfrac{I^0(K_0, E^S)}{p}$.

I emphasize that the only substantial ways in which this equilibrium differs from a private-ownership Arrow-Debreu equilibrium are that the state carries out all the investment, and the labor-supply decisions do not comprise a *Nash* equilibrium for workers, but an additive Kantian equilibrium. In determining their supplies of labor, workers do not ask whether supplying an extra day's labor generates an after-tax wage that compensates for the extra disutility of labor, assuming all other workers' supplies of labor remain fixed, but rather, whether the extra day's income compensates for the extra day's work, *assuming all others expend an extra day's labor as well,* which will increase substantially the value of the demogrant.

The consequence of this *single* change in optimizing behavior is the following:

Proposition 4.1 *Let* $(p, w, r, K_0, E^1, ..., E^n, x^0, ..., x^n)$ *be a Walras-Kant equilibrium at any tax rate* $t \in [0,1]$. *The equilibrium is Pareto efficient.*

One can call this proposition the first welfare theorem of market socialism. It states that the equilibrium is Pareto efficient *at any income tax rate.* In particular, society, by democratic choice of the tax rate, can achieve any degree of income equality it desires – even perfect income equality, when $t = 1$ -- with no sacrifice in efficiency.

To be precise, I must explain what Pareto efficiency means here. Think of the state's having the utility function $u^0(x) = x$; the state simply cares about the amount of the investment good it acquires for the 'next period,' placed in quotes because the model has only one period. An allocation is Pareto efficient if there is no other feasible allocation that renders some agent better off, and none worse off, where the state is included as an agent. Now since there is no future in this model, we cannot assert that the state's investment is that which trades off efficiently present and future consumption of workers – because there is no future in the model. Most economists believe that the Soviet Union and China invested too high a fraction of the national product – these states were not (probably) trading off present and future consumption of their citizens properly. That, too, can happen in the present model. But to deal with this problem, one would have to articulate a multi-period model, and I have elected here to keep things simple by truncating the future.

What is the trick that generates what appears to be an astounding result – that there is no trade-off between efficiency and equality in the model? The dead-weight loss of income taxation has been, apparently, entirely nullified by Kantian optimization in the labor supply decision. Without giving a full proof of Proposition 3.1, I can indicate how this occurs.

Let's ask how a worker sees his income as varying if he *and all other workers* increase their labor supply by a small amount. We must calculate:

$$\frac{d}{d\rho}\bigg|_{\rho=0} I^i(E^1 + \rho, \ldots, E^n + \rho); \tag{4.5}$$

expanding this derivative shows it is equal to:

$$(1-t)w+(1-t)\theta^i \Pi_2(K_0,E^S)n+\frac{t}{n}(pG_2(K_0,E^S)n-\theta^0\Pi_2(K_0,E^S)n) \quad (4.6)$$

where G_2 and Π_2 are the derivatives of G and Π with respect to their second

component.   Now at equilibrium, we have:

$$pG_2 = w \text{ and } \Pi_2 = 0 \; ; \qquad\qquad (4.7)$$

both statements are directly the consequence of the firm's having maximized profits.

Therefore the mathematical expression in (4.6) reduces to:

$$(1-t)w+0+\frac{t}{n}wn+0 = w \; . \quad (4.8)$$

That is, in contemplating the counterfactual stipulated in the definition of additive

Kantian optimization, the worker computes that an extra's days work will supply her *not*

with the after wage $(1-t)w$, but with the *gross* wage, w.   This is because what the

worker loses in her personal tax bite, she regains in the increased value of the demogrant.

So the appropriate optimization condition is that the worker equate her gross wage to the

marginal rate of substitution between labor and consumption.   But this is the condition

for Pareto efficiency!

In contrast, the worker who Nash-optimizes assumes that if she increases her

labor supply by a small amount, all other workers stand pat, and this means the worker's

optimality condition is equality of the after –tax wage $(1-t)w$ and the marginal rate of

substitution between labor and consumption. Thus is generated the dead-weight loss of

taxation in the standard Arrow-Debreu model.

Do these equilibria exist?   Under standard conditions, they do:

Proposition 4.2 *If the production function G is concave and satisfies standard conditions,*

and each utility function is strictly concave and the demand for the commodity is normal,

then a Walras-Kant equilibrium exists for all tax rates $0 \leq t < 1$.[4]

---

[4] The 'standard conditions' on *G* are the Inada conditions and homotheticity; a
consumption good is *normal* if an increase in income generates an increase in the good's
consumption.   These conditions are sufficient for equilibrium; they may not be
necessary.

Furthermore, the propositions remain true with many firms and many goods. What happens if private citizens can also invest in the firm(s)?   To preserve the first welfare theorem, we would have to require that the vector of investments also be an additive Kantian equilibrium. This strikes me as being a less plausible assumption than the labor-supply assumption[5].

5. Discussion

The first model of market socialism is credited to Oscar Lange and Fred Taylor (1938).   I have reviewed their model elsewhere (Roemer [1994]), and the discussion it generated, importantly the critiques made by F. Hayek (19xx).   I shall not do so again here, except to say that the model's principal feature was state control of investment. Lange and Taylor also proposed a kind of tâtonnement process, in which the central planning board imitated convergence of prices to what they imagined markets did in reality.   That part of their model has been discredited by later work of general-equilibrium theorists, who showed that the tâtonnement process does not in general converge to the equilibrium, even when one exists.[6]   There have been a plethora of market-socialist models where the state or other public institutions carry out investment; the most recent contribution is by G. Corneo (in press).   The focus of all these models has been upon equalizing the distribution of capital income in society: the state spends its profit income either on investment or public goods and services or demogrants. But these models have had relatively little to say about labor income, which in advanced economies comprises at least 60% of national income.   Informally, we can imagine that the state would invest substantially more in compensating children from disadvantaged households to build up their skills, and it may also engage in redistributive taxation. But these facets have not been formalized in novel ways.

One must also mention the theory of labor management, in which workers own firms, and maximize valued added per worker.   The most rigorous presentation is due to Jacques Drèze (1989), who shows that the equilibria in a labor-managed economy are

---

[5] The results exposited in this section, in a somewhat more complicated model, are given in Roemer (2017).

[6] See Sonnenschein (1972), Debreu (1974), Mantel (1974).

isomorphic to Arrow-Debreu equilibria (without taxation), and are hence Pareto efficient. There is, however, no formal discussion of the degree of freedom that exists (or otherwise) with respect to income distribution.

The novelty of the present approach is that it embeds a formal model of worker cooperation into a market economy, something that has not been done earlier because no such model of cooperation existed. Of course, forming a worker-owned firm or agreeing to have the state manage investment can be thought of as a form of social cooperation, and so my latest statement may seem narrow-minded. What I intend to say is that cooperative behavior in decision making has not been formally injected into market socialist models heretofore, and it would seem to be an advance to do so, if we take seriously the idea that *socialist ethos* is something quite different from capitalist ethos, and deserves a formalization.

Finally, I wish to address the question that most readers will undoubtedly have about the proposed market-socialist model – and indeed, about the rather complicated forms of behavior stipulated in multiplicative and additive Kantian optimization more generally. It might be useful to say that Nash optimization was probably viewed with some trepidation as a characterization of human behavior when Nash introduced it as a graduate student in mathematics in 1950. Von Neumann apparently rejected it out of hand when Nash explained it to him. We still lack an adequate theory of how players converge to a Nash equilibrium; often, a sequence of iterated best responses does converge to an equilibrium, but the weakness of this mathematical fact as an *explanation* of convergence is that at each step in the process, the assumption that players make that other players' actions are fixed is belied[7]. Yet we do seem to have ample evidence that many stable situations appear to be describable as Nash equilibria of social interactions that can be described as games. And as I've said, even behavioral economists, who challenge the most classical interpretation of these equilibria do not depart from describing them as Nash equilibria of games, albeit with exotic preferences.

I think three pre-requisites are necessary for believing that a society of workers could actually optimize their labor-supply decisions in the manner postulated by additive

---

[7] In like manner, a sequence of iterated best responses also often converges to a Kantian equilibrium. See Roemer (in press, chapter xx).

Kantian optimization: desire, understanding and trust.   Workers must *desire* to cooperate, they must see themselves as part of a solidaristic venture, in which cooperation could further their interests.   They must *understand* that if they act in concert as the Kantian optimization process stipulates, the results will be good, in the sense that one can separate the issue of efficiency from that of distribution. Citizens can decide on how much redistribution they believe is warranted, due the unfairness of the distribution of endowments that will necessarily exist, without having to compromise because too much redistribution might be inefficient.   Thirdly, each must *trust* that if he optimizes in the Kantian manner so will most others: he will not be taken advantage of by Nash players.

We have many examples of Kantian behavior in history, and desire, understanding and trust surely characterized those occasions.   Although the labor-supply decision may seem more complex, in fact it is not cognitively more difficult than Nash optimization. Each person need only know his own preferences, as in Nash optimization; and instead of standing pat when his marginal rate of substitution is equal to the net wage, he stands pat when it is equal to the gross wage.

Granted, education will be required.   At present, in the United States, many people do not seem to understand that their taxes are used for *any* useful purpose, let alone returned to them in transfer payments, services and public goods.   Europeans, being less state-phobic, will be more receptive to understanding the virtues of Kantian optimization.   This survey, however, has not attempted to address how we get there from here: rather, its more modest aim has been to offer a design that can motivate our imagination about a possible future.
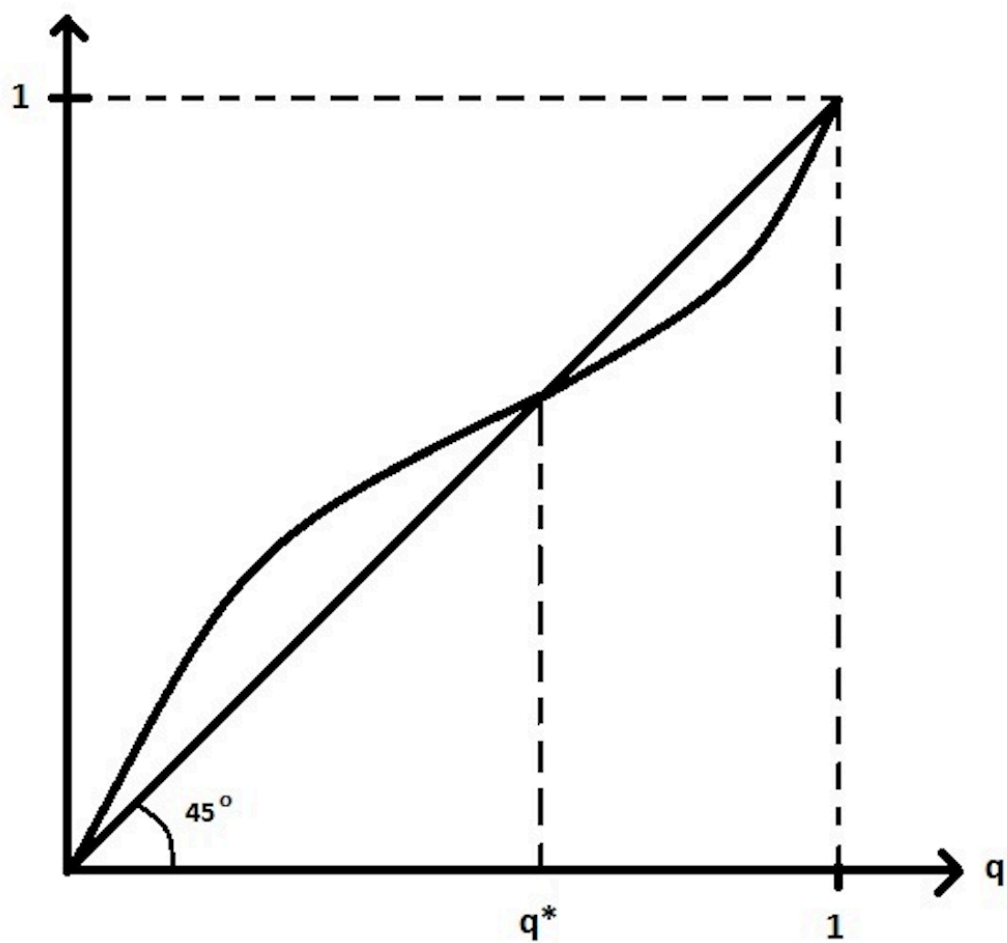
Figure 1 The curve is the distribution function of thresholds q. The stable equilibrium is that fraction q^* of the population recycles.

References

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

Barbera, S. and M. Jackson 2016. "A model of protests, revolution, and information,"

Cohen, G.A. 2009. *Why not socialism?* Princeton university Press

Corneo, G . 2018. *Is capitalism obsolete?* Harvard Univ. Press

Debreu, G. 1974. "Excess demand functions," *J. Math. Econ.* 1, 15-21

Drèze,J. 1989. *Labour management, contracts and capital markets: A general equilibrium approach,* Oxford Univ. Press

Dunbar, R.I.M. 2009. "The social brain hypothesis and its implications for social evolution," *Annals of human biology* 36, 562-572

Elster, J. 1981. "States that are essentially by-products," *Social science information* 20, 431-473

Elster, J. 1989. "Social norms and economic theory," *J. Econ. Perspectives* 3, 99-117

Elster, J. in press. "On seeing and being seen," *Social choice and welfare*

Hayek, F.

Kandori, M. 1992. "Social norms and community enforcement," *Review of economic studies* 59, 63-80

Lange, O. and F. Taylor 1938. *On the economic theory of socialism*, Univ. of Minnesota Press

Mantel, R. 1974. "On the characterization of aggregate excess-demand," *J. Econ. Theory* 7, 348-353

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

Roemer, J. 1994. *A future for socialism,* Harvard Univ. Press

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

Roemer, J. 2017. "A design for market socialism," Cowles Foundation Discussion Paper xx, Yale University

Roemer, J. In press. *How we do and could cooperate: A Kantian explanation,* Yale University Press

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

Sonnenschein, H. 1972. "Market excess demand functions," *Econometrica* 40, 549-563

Tomasello, M. 2014. *A natural history of human thinking*, Harvard UP

Tomasello, M. 2016. *A natural history of human morality,* Harvard UP