

*John E. Roemer*

## Thoughts on G. A. Cohen's Final Testament\*

*Abstract:* I present briefly G. A. Cohen's theory of distributive justice, discuss the relationship that I think he believed held between human nature and justice, and offer thoughts on the feasibility of Cohenesque justice, or Cohenesque socialism. I introduce the idea of Kantian equilibrium, as a way of explaining how people cooperate. Expanding the domain of activities in which humans cooperate will, I believe, go a long way towards achieving Cohenesque socialism, and the history of human society suggests it is feasible to do so.

I will discuss three topics. First, I will provide a brief review of what I take to be Jerry Cohen's theory of justice, which derives from his essay, published in 2009, the year he died, entitled *Why not Socialism?* Second, I'll discuss what I believe Jerry's view was concerning the relationship between human nature and justice. Third, I will offer some views on the feasibility of Cohenesque justice. The last part will be social science, or rather crystal-ball-gazing social science, not philosophy.

### 1.

In *Why not Socialism?* (Cohen 2009) Jerry wrote that justice comprises socialist equality of opportunity. Equality of opportunity comes in three versions, which he called bourgeois, left-liberal, and socialist. Each version eliminates inequality that is due to certain institutions, or property in certain kinds of asset. Bourgeois equality of opportunity eliminates coercive labor relations, the kind characteristic of slavery and feudalism. It is the kind of opportunity-equality which exists in an ideal form of capitalism, where every individual has the legal right to enter into contracts with any other, and invidious discrimination does not exist. Left-liberal equality of opportunity eliminates in addition *socially* generated inequalities due to the families into which individuals are born—the unequal advantages from which people benefit due to differential family wealth, parental education, parental social connections, and so on. Socialist equality of opportunity eliminates in addition inequalities of condition due to inborn traits

---

\* Prepared for American Philosophical Association meetings, Baltimore, December 2013.

of individuals—intelligence and talent. What kind of inequality remains in a regime with socialist equality of opportunity? Inequality due to choice. It may not be easy to decide when inequalities that are due to variations in intelligence and talents, and due to families which raise children, have been eliminated. It is not necessary to do so from a conceptual viewpoint. It may be difficult to decide when inequalities are due to something we would call autonomous and premeditated choice, but that is another question. The theory leaves open the possibility that there is *no* realm for choice, and hence that all inequalities of condition must be eliminated for justice. Thus, in principle, the case of the non-existence of free will is a special case of the theory, which requires neither that free will exist nor that it does not exist. Jerry was a compatibilist, and believed that some actions of individuals can properly be attributed to choice, even given a biological/physical view of the micro-foundations of human action.

There is another kind of inequality, which Jerry did not discuss, but which I believe would remain in a regime of socialist equality of opportunity. Parents pass down to their children wealth, education, aspirations, social connections, genes, and preferences or values. I think that parents may rightfully influence their children's preferences and values in ways which will lead to inequality in material condition among the generation of children when they become adults. To the extent that parents pass down to children their views on the right way to live, and the genesis of these preferences is not polluted by cognitive dissonance, the influence is legitimate, in my view.

Inequalities would remain, even under socialist equality of opportunity, not only due to legitimate choice, but due to the playing out of option luck, and to careless or sloppy choice. Both kinds of inequality are legitimate at the bar of socialist equality of opportunity. Of course there would be inequalities not due either to option luck or sloppy choice, but to deterministic consequences of different, premeditated choices. The existence of large inequalities of this sort, however, may preclude the maintenance of an ethos of fraternity or community among people. Jerry believed that fraternity was a necessary condition of the good society—he was unsure whether fraternity should be included in the definition of justice or not. In either case, he wrote that large inequalities should be forbidden, even if they were legitimately generated under a system of socialist equality of opportunity, because they would compromise fraternity. So his all-considered view was that justice, or at least the good society, comprised socialist equality of opportunity with the size of inequalities generated therein constrained by the necessity of preserving fraternity or community.

One might ask what the Cohenesque currency in which inequality is measured is. Jerry never attempted to give a precise answer to this question. In his article 'On the Currency of Egalitarian Justice' (Cohen 1989), he gave examples showing that that currency included aspects of both resources and welfare. A man is able to move his arms, and do anything with them that normal people can do, but only with great pain. Justice requires that we provide him with pain-relieving medicine. His deficit is entirely one of welfare, Jerry says, because his arms are as adequate a *resource* as the arms of normal people. Thus, the currency must include, to some extent, welfare. On the other hand, the cheerful Tiny Tim,

although happier than most people, cannot walk: justice requires that he be provided with a wheel chair. His deficit is one only of a resource, not of welfare. So the currency must include a measure of resources, to some extent. I do not think that Jerry said much more about the currency of justice than this.

Some may ask: if voluntary exchanges are permitted between people after inequalities due to family background and internal talents have been eliminated, why cannot capitalist inequality be generated, through Nozickian capitalist acts between consenting adults (Nozick 1974)? If so, the constraints demanded by insuring community would indeed be extremely important. Jerry, in his earlier work, made several replies to this query. Perhaps the most important one was to challenge Nozick's neo-Lockean proviso, that appropriations of unowned resources in the natural world by individuals are morally justified as long as the owner leaves others at least as well off as they would have been had the resource remained unowned. Thus Frank, who uniquely knows how to use the sand on the beach to make antibiotics, can appropriate the beach, which had been used by the community in common for swimming, as long as the antibiotics he makes from the sand and sells to people to cure their illnesses leave them at least as well off as they were when they could swim from the beach but died of diseases for lack of antibiotics. Frank can even, according to Nozick, charge the monopoly price for his antibiotics as long as the proviso continues to hold. Jerry responded that the benchmark of the proviso was arbitrary: Why not say that property in the natural world that is not owned by any individual is owned by all? If this were so, Frank would have to *bargain* with the community over the price of the antibiotic before ownership of the beach was transferred to him. Presumably, the community could extract a large part of the economic surplus created by the invention of the antibiotic, through bargaining. This challenge to Nozick's edifice does not challenge self-ownership as such. Jerry considered it of great importance to destroy Nozick's argument while not denying the premise of self-ownership, because he believed that doing so could recruit many to an anti-capitalist view, many who would be reluctant to deny the thesis of self-ownership.

Jerry's conception of socialism, as socialist equality of opportunity with constrained inequality of outcomes to preserve community, was an ethical conception. The issue of what institutions could be used by a society guided by the socialist ethos to achieve socialism is another question—whether those institutions require state ownership of productive assets, worker-ownership, or whatever. He was at pains to distinguish rules of social regulation, which would specify institutional structure, from theories of justice. I am sure he would agree with me that the societies most closely approximating socialist justice today are the Nordic countries, where private property in productive assets is extensive. In particular, these societies are closer to socialism than were the Soviet-type societies in which private property and markets were in large part eliminated or severely constrained.

## 2.

My second topic is the relationship between human nature and justice. Jerry was an optimist about human nature: he wrote in *Why not Socialism?* that the main constraint to achieving socialism is not human nature, but the problem of design. We know, he said, how to construct institutions that organize an economy, relying on the motivations of greed and fear—namely, markets with private property in productive assets—but we don't know how to organize an economy based on socialist principles. Those principles are, as I said, socialist equality of opportunity and fraternity. The ethos of capitalism is self-interest and greed; the ethos of socialism is one in which 'people care about, and where necessary and possible, care for, one another, and, too, care that they care about one another'. This kind of 'caring about' and 'caring for' is a delicate ethos, that cannot be preserved in the face of large inequalities of condition, or so he claimed. That impossibility must be due to human nature. So if justice is so defined—in particular, including the proviso of limiting legitimately generated inequalities—then that's because human nature is what it is.

This raises an issue. In his penultimate book, published in 2008, *Rescuing Justice and Equality*, Jerry argued that moral theories, and justice in particular, do not depend upon facts about the world. Suppose you propose a theory of justice, call it *B*, that is predicated on facts about the world. Then what really guides you, Jerry claims, is a more general theory, call it *A*, that says, 'If the facts about the world are *x*, then justice requires *B*, and if they are *y*, then justice requires *B*'. Theory *A* is independent of the facts about the world. Jerry conjectured that whenever a theory of justice appears to depend on facts, there really exists a prior more general theory, some steps back, that is fact-independent, from which the fact-dependent theory is deduced. And that prior theory is the general theory of justice.

Now including the constraint to limit inequalities in order to preserve community is only necessary because, and if, human nature is incapable of maintaining the communitarian ethos in the face of large inequalities, which are generated by a regime of socialist equality of opportunity. Jerry wrote,

“Although inequalities [due to socialist equality of opportunity] are not condemned by justice, they are nevertheless repugnant to socialists when they obtain on a sufficiently large scale, because they then contradict community: community is put under strain when large inequalities obtain.” (Cohen 2009, 11)

Jerry gives the example of a man, call him Jack, who has accumulated many material comforts, due to actions which are justified under socialist equality of opportunity. One day Jack's fancy car is in the shop, and he must take the bus to work, mixing with others who have elected not to spend their earnings on cars, or more precisely, not to take jobs which generate a sufficient income to purchase fancy cars, and who must take the bus every day. These others cannot commiserate with Jack if he complains about the inconvenience of taking the bus—community is impossible, says Jerry, given the material inequalities of

condition among these folks. But surely that is only because the impecunious ones who do not own cars have not fully taken on board the justice of socialist equality of opportunity: they chose not to earn such high salaries, because of their preferences, let us say, for more leisure, or for less remunerating kinds of work. They cannot commiserate with Jack due to a powerful dimension of human nature, which prevents sympathy in the face of material inequality. Therefore, Jerry's proposed definition of justice is not, indeed, independent of facts about the world. He really must possess a more general theory which says, 'If human nature renders community impossible when material inequalities are too great, then socialist equality of opportunity must be constrained to enable community, and if human nature does not render community impossible in the face of large inequalities, then socialist equality of opportunity *tout court* is justice'.

Indeed, Jerry criticized Rawls for the following inconsistency in his argument. One requisite of the just society, for Rawls, is that it be well-ordered: that its members endorse the theory of justice, in particular, the difference principle (Rawls 1971). But if everyone endorses the difference principle, should they not choose, in their personal lives, to take those actions that would render the least well off as well off as possible? If they did, the incentive problem would cease to exist. A talented person would not desist from exercising her talents if her income were highly taxed, assuming that those tax revenues would increase the welfare or primary goods allocated to the worst off. Now is Jerry perhaps guilty of a similar error? If everyone in society agrees with the thesis that socialist equality of opportunity comprises justice, and therefore the inequalities that come with it are just, why should those inequalities compromise community? Are the impecunious bus riders who resent Jack not exhibiting a lack of agreement with the theory of justice that governs their society? And if Jerry would respond that they just can't help it, the resentment comes with their human nature, should he not allow Rawls a similar retort on the incentive issue?

The two cases are not quite isomorphic. Jerry claimed that Rawls's theory was *inconsistent*, because well-orderedness requires that people take decisions in their personal lives which maximize the well-being of the worst off class, and this would imply full equality of condition, rather than the inequalities that the difference principle accepts. Perhaps the difference principle should be considered a rule of social regulation—the closest thing to equality that society can access, given a somewhat self-interested human nature. I would say that Jerry's own fundamental principle of justice was socialist equality of opportunity, unconstrained. That may be *undesirable* because of human nature that cannot maintain community in the face of large material inequalities. This does not make Jerry's theory inconsistent, as long as he does not insist, as did Rawls, on well-orderedness. But it would mean that the constraint that disallows large material inequalities that may prevail under socialist equality of opportunity is only necessary because of a weakness in human nature which prevents persons from living fraternally in a state of justice. It is also possible that justice should *include* fraternity as a condition, but Jerry was reluctant to insist upon this. If fraternity is a condition of justice, then full justice is either impossible, or

some compromise between fraternity and socialist equality of opportunity must be accepted if human nature requires it.

A propos the socialist ethos, which became increasingly important for Jerry as a prerequisite of socialism, and which he attempted to illuminate with the example of the camping trip in *Why not Socialism?*, I reproduce a quotation from a speech made by Per Albin Hansson, in 1928, the year he became chair of the social-democratic party SAP in Sweden.

“The basis of the home is community and togetherness. The good home does not recognize any privileged or neglected members, nor any favorite or stepchildren. In the good home there is equality, consideration, cooperation, and helpfulness. Applied to the great people’s and citizens’ home this would mean the breaking down of all the social and economic barriers that now separate citizens into the privileged and the neglected, into the rulers and the dependents, into the rich and the poor, the propertied and the impoverished, the plunderers and the plundered. Swedish society is not yet the people’s home. There is a formal equality, equality of political rights, but from a social perspective, the class society remains and from an economic perspective the dictatorship of the few prevails.” (Berman 2011, 245)

I am sure Jerry would have cited Per Hansson in *Why not Socialism?* had he known of this speech.

Let me comment, tangentially, on the relationship between what I take to have been Jerry’s final conception of justice and his earlier concern with Marxist exploitation, which he argued, despite some of Marx’s claims to the contrary, must be interpreted as a theory of injustice. Exploitation is formally defined as an inequality in labor accounts: workers are exploited when the labor embodied in the goods that they can purchase with their wages is less than the labor time they expend in productive activity. The difference in these amounts of labor, so-called surplus labor, redounds to the capitalist, embodied in the product of workers that she owns and sells at a profit. Workers agree to sell their labor power to capitalists in order to gain access to the means of production, owned by capitalists, which increases the productivity of their labor, engendering wages higher than then could earn without access to those means. Their alternative, let’s say, is to scrounge for a living on unproductive plots of land that are accessible to all. Thus exploitation derives from two facts: that the means of production are concentrated in ownership, not equally distributed, and that there is a scarcity of means of production relative to the labor available for employment. The latter fact bids the real wage down to a minimum, which, in a competitive setting, would be only slightly greater than the real wage workers could earn on the freely accessible low-productivity plots.

Now the accumulation of capital in the hands of a small class can easily be due to relations which are acceptable under bourgeois equality of opportunity: those who become capitalists are those with rich families, good educations, inborn talent, and social connections. Indeed, actually existing capitalism is replete with episodes of accumulation that are not acceptable even under bourgeois

equality of opportunity—robbery, plunder, enclosure, and so on. Even without the latter, however, exploitation would be *in large part* due to inequalities in the ownership of productive assets whose genesis would be illegal under a regime of socialist equality of opportunity. Therefore, Jerry's final view of justice is a generalization of his earlier concern with exploitation. It is a more fundamental theory than the theory of justice as the absence of exploitation.

Jerry pointed out that Marx's theory of historical materialism drew its power from a confluence of facts characterizing nineteenth century capitalism: that the working class had the *power* to bring down capitalism, that they were the *poor*, and that they were the *exploited*, in the sense of being those who were unjustly treated. In present-day capitalism, at least in the advanced societies, it is arguable that the poor, the exploited, and those who have the requisite power are three different groups. (The poorest are those who cannot find work, and don't have the good luck of being exploited, and those who have the power may not be the industrial working class, but highly paid and highly skilled workers.) And furthermore, the identification of the very wealthy with the capitalist class no longer clearly holds. A century ago, households in the top 0.1% of the income distribution earned almost all of their income from capital, in the form of dividends, interest, and rents. Today, a substantial fraction of the income of the top 0.1% is labor income—salaries of CEOs, bond traders, star athletes, movie stars, and some highly paid professionals. Some, who would protest that the incomes of these people should really be classified as part of profits, ignore the definitions. It is absolutely clear that star athletes, movie stars, bond traders and even CEOs compete on highly competitive labor markets; and surgeons and lawyers either compete on such markets, or are partners in worker-owned firms. Their incomes are rightly classified as labor income. What must be revised is not the definition of labor income, but the view that the very wealthy under capitalism comprise the capitalist class, and the identification of the exploited with the unjustly treated. The high fliers I have just mentioned are workers in the sense that they sell their labor on competitive markets, and those who do not own their own firms, like lawyers and surgeons, are, I claim, still exploited in the Marxian sense—they pay a rent to the owners of the firms for which they work in order to have access to those firms' technologies. (This must be true of star athletes or movie stars: otherwise, why would the sports-team owners and film studios hire them? And how much would a movie star's charisma be worth without the movie studio? So it's well worth it to him to sell his labor to the studio, even though he does not receive his entire product.) But it is extremely difficult to imagine that these high fliers, exploited in the Marxian sense though they may be, are unjustly treated. They benefit massively from the possession of assets which would garner much smaller rewards were socialist equality of opportunity in place.

## 3.

The third topic I promised to discuss is the feasibility of Cohenesque socialism. Jerry wrote:

“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, 57)

Although there exist two cases—either human nature does or does not preclude the possibility of socialism—the interesting case to consider is that it does not, for that is the only case where those who aim to achieve socialism are not deceiving themselves. Still, one must say something about human nature, for surely one of the greatest arguments of the Right is that human nature is inherently selfish, thereby rendering socialism impossible. There is now much work in evolutionary biology on the cooperative nature of *homo sapiens*: our species is not unique in this regard, but surely cooperation is far more developed in us than in the other species of higher mammal. I would remind you of three pieces of evidence. The clear tendency in economic development is towards *urbanization*, a condition in which millions of human beings live together, cheek by jowl, almost without killing each other. The *decline in violence*, which has come with economic development, is also quite stunning. Even in the twentieth century, with its two world wars, the fraction of the total population killed by violence was substantially less than the corresponding fraction in the pre-agricultural period of hunter-gatherer society. Steven Pinker, in *The Better Angels of Our Nature* (Pinker 2011), offers a convincing compendium of the decline of human violence. Thirdly, the *economic role of the state*, in advanced economies has become huge. Roughly one-half of the gross domestic product in northern European countries is appropriated by the state and spent on public goods, state-provided private goods (like education), social insurance, and redistributive transfers. This represents a substantial collectivization of the economic product, although it has not come about through the mechanism that Marx pointed to, collective ownership of productive assets, but rather through taxation.

The decline in violence has been occurring for centuries, if not millennia: but urbanization and the large state share in the economy are relatively recent phenomena. Urbanization occurs only after agriculture becomes sufficiently productive that only a small fraction of workers must remain in agriculture to feed society. And the large state share in the economy is a development of the last century. Given the recency of these developments, it is not utopian to believe that further strides towards socialist society will occur in the next half millennium.

To be precise, one must distinguish between Jerry's communitarian ethos and *cooperation*, and so the evidence of cooperation I've just given does not necessarily imply development of a socialist ethos. People may engage in cooperation even if they are purely self-interested: if they can cooperate, each can achieve higher welfare for herself than otherwise. Urbanization, for instance, concentrates highly skilled people geographically, making possible economies of agglomeration (Silicon Valley, Wall Street, Route 128 around Boston). Non-cooperative behavior often destroys welfare: think of the tragedy of the commons, where the fishery is destroyed because fishers do not regulate their self-interested activity. The political scientist Elinor Ostrom observed that small communities very often learn to regulate fishing, and similar activities, and avoid the non-cooperative Nash equilibrium which gives rise to the tragedy of the commons (Ostrom 1990). Presumably—at least, so Ostrom assumed—the cooperation is successful not because fishers possess the Cohenesque ethos, but because constraining their individual actions renders them better off as individuals.

So the evolution of cooperation that I've pointed to is not equivalent to possession of socialist ethos. The welfare state, for example, came about not primarily due to the concern of the haves for the have-nots, but rather as a system of social insurance, which was in the self-interest of the vast majority. It is rational for each of us, as individuals, to purchase insurance against calamities of illness and unemployment. For reasons of moral hazard and adverse selection, it may be unprofitable for private enterprise to offer such insurance. If there is a means through which choice can be coordinated, such as a democratic election, self-interested individuals may choose to design state insurance schemes, which are, in large part, what comprise the welfare state.

I began the query into socialism's feasibility by pointing to evidence of the recent cooperative achievements of human beings, and then took a step back, observing that cooperation is not necessarily motivated by a socialist ethos, but makes good, self-interested sense. My conjecture, however, is that cooperative achievement fosters socialist ethos, because it shows people that all can advance together. *Learning* that we all can advance together, but not alone, is, I think necessary, although perhaps insufficient, for encouraging an ethos where 'people care about, and where necessary and possible, care for, one another, and, too, care that they care about one another'.

It is difficult to predict how human psychology will evolve as the period of material privation fades into history. We are living on the cusp of this historical event. Perhaps the best evidence for it is the very rapid advance in life expectancy in the last century. Global life expectancy increased from 31 years in 1900 to 67 years in 2010; in the United States, the numbers are 49 years in 1900 and 77.5 years in 2003. I conjecture that the decline in violence, over the long period, that Steven Pinker measures, is due to reduced material deprivation—that is, less intense competition for the means to survival—which also explains the very rapid increase in life expectancy. Due to our relatively recent experience with this new abundance, we have yet to learn to control our appetites. Those of us who succeed financially are almost all subject to the *nouveau-riche* penchant of over-consumption. But I believe that that control will develop fairly

quickly, and we will cash out our new productive capacity in more leisure time. Indeed, if we amortize the length of the present work-year in Germany over a fifty-week year, then Germans at present work 28 hours a week, probably less than half what they worked a century ago. The average annual work time in the US, amortized over a fifty-week year, is among the highest in the advanced countries, at 36 hours per week.

If the ending of material privation, on a global scale, reduces violence, and the habits necessary for self-preservation that evolved with it, we can conjecture that it will eventually induce an increase in the concern for others, and the socialist ethos that Jerry defined. Certainly, there are many, even today, who exhibit a strong concern for others beyond their family members: such people often enter the teaching and caring professions or, more generally, choose their occupations with an eye to helping humanity. One cannot underestimate the effectiveness of capitalism in instilling an ethos of individualism and self-interest, an effectiveness that, I claim, is facilitated by conditions of material privation and the consequent urgency of focusing upon survival.

#### 4.

So much for my expounding of Jerry's claim that the key problem for socialism is not the innate selfishness of humans. I now turn to the design problem. I have argued against the view that human nature precludes the development of a socialist ethos. But the design problem remains: if markets harness the emotions of greed and fear, and indeed generate greed and fear, as Jerry evidently believed, what institutions can coordinate economic activity, which are consistent with, or harness, a socialist ethos?

I believe that we will continue to need and use the market mechanism for more centuries than I can contemplate, because of its ability to coordinate economic activity in large societies. I do not, however, agree with Jerry that markets either require or necessarily induce greed and fear in those who trade on them. It is difficult to disentangle the role of *markets* and of *material privation* in generating greed and fear. All of the readers of this essay use markets pervasively, yet very few of us, I wager, feel the emotions of greed and fear in our economic transactions. This is not mainly because we have contemplated philosophy, but because we do not face scarcity or insecurity. If scarcity and insecurity increasingly become things of the past, as I have argued they will, our own lack of greed and fear in the presence of market transactions will become universal.

Indeed, one can argue that the Nordic countries exhibit the truth of this prediction. Security is high with the Nordic welfare state, and fear has probably largely been eliminated. And the Nordic flattening of wage and salary differentials, which began as a strategy for weeding out inefficient firms, proposed by Swedish economists in the 1950s, has been largely maintained, and has not, evidently, diminished labor productivity, indicating that greed is not terribly active.

Is the design problem, then, not as serious as Jerry proposed? Could we not take the Nordic institutional design as the template? We could do far worse—and by ‘we’, I mean global human society. But I believe we may be able to do far better.

Let me explain how. I am going to describe a ‘third way’, but it is not one of the usual kind that go by that name. I do not focus on the ownership of productive assets—from the Nordic case, it seems to me that collective control of national income requires neither collective ownership of productive assets nor worker-ownership of firms. I will assume something less than socialist ethos: namely, that people understand the value of cooperation, if only because it fosters their own well-being. I propose there is a vitally important component to economic behavior, which is conceptually separate from individual preference, and which, in current economic theory, is not recognized as something which is subject to possible variation, and therefore is not mentioned as one of components of economic coordination. This is the *optimization protocol* that people use in economic calculation. By this phrase, I mean *the way that people optimize*.

Let me be more specific. The optimization protocol that economists assume is invariant across rational persons, and therefore need not be specified as a token of a more general type, is what I will call *autarkic optimization*. This is formalized in the concept of the Nash equilibrium of a game. A Nash equilibrium is a vector of actions or strategies, one for each person in the game, such that no individual would improve his own welfare by altering his action, assuming that all others’ actions remain fixed. It is clear that, *if* everyone optimizes in this fashion, and if, somehow, a collection of players finds themselves at a vector of actions which is a Nash equilibrium, then that vector of actions is indeed stable—it is indeed an *equilibrium*. For no one will have the incentive to change his actions assuming no others change theirs, given the way each is thinking. In the past 60 years, since John Nash proposed the concept, what is now called Nash equilibrium has become the ubiquitous concept of equilibrium in economic theory. Indeed, the competitive equilibrium of an economy, which was conceptualized in the 1870s by Leon Walras, can be viewed as a very special case of Nash equilibrium.<sup>1</sup> Earlier economists, including Auguste Cournot in the 1830s and Harold Hotelling in the 1920s proposed equilibrium concepts for special economic games that are also special cases of Nash equilibrium. So, although the idea was not brand new when Nash proposed it, he formalized what was in the air, made it into what Joseph Schumpeter aptly called ‘tooled knowledge’. Once the concept had entered the lexicon, with John Nash’s 27-page Ph.D. dissertation in mathematics at Princeton University, it rapidly colonized economic theory, and indeed, much of social science.

Nash equilibrium is known as a *non-cooperative* equilibrium concept, which is a synonym for my adjective ‘autarkic’. The non-cooperativeness, or autarky, is instantiated in each players’ contemplating that *he alone* might change his action, while assuming the others are inert. This non-cooperativeness is at the

---

<sup>1</sup> Its ‘very’ special property is that, under suitable assumptions, the competitive equilibrium is Pareto efficient (the ‘first theorem of welfare economics’), contrary to the virtually generic inefficiency of Nash equilibrium in games. See below.

root of the tragedy of the commons and the prisoners' dilemma. Almost always, Nash equilibria are Pareto inefficient: that is, in most games, there exist vectors of actions at which all players would be better off than they are at the Nash equilibrium action profile of the game. But those Pareto-dominating vectors of actions are inaccessible, given autarkic reasoning. They cannot be achieved by individuals' moving singly from the Nash equilibrium, but only by their moving together. And furthermore, these new Pareto efficient vectors of actions are usually not themselves Nash equilibria. So autarkic—one might say opportunistic—behavior will unravel them.

Indeed, the tragedy of the commons is only a special case of the general problem of externalities in production. As long as the action of each individual has knock-on effects for the welfare of others, externalities exist, and when they do, Nash equilibria are Pareto inefficient. In the case of the fishing economy, those externalities are negative: when I increase my fishing time on the lake, I render the lake a less productive asset for others. But this is not something that I take into account in autarkic optimization. When these externalities exist, Nash equilibria are inefficient.

Suppose, instead of a fishing economy, we are in a hunting economy. All of us hunters go into the jungle to search for big game, and let us suppose that we have agreed to divide the spoils of the hunt equally among us. Anthropologists believe this was a traditional allocation rule in hunting societies. In the Nash equilibrium of the game where each hunter's strategy is his hunting effort, each hunts *too little*. It is in my selfish interest to find a bush to nap under, if others are hunting vigorously: doing so will not reduce by very much the possibility of our bagging a buffalo, and in that case, I will get my per capita share anyway. So hunters offer *too little* effort in the Nash equilibrium of the hunting game, from the social point of view, while fishers offer *too much* effort in the Nash equilibrium of the fishing game, from the social point of view. That's because the allocation rules are different: recall that in the fishing game, each keeps his catch, whereas in the hunting game, the catch is shared equally. Hence the mathematical structure of the fishing economy and the hunting economy are different. Nash equilibria in both economies are Pareto inefficient.

Now I propose a different optimization protocol in the fishing game, where each fisher keeps his catch, and there are negative congestion externalities on the lake. Suppose I am contemplating increasing my fishing time by 10%—the extra fish I catch would be, I compute, well worth the disutility of the extra labor I would expend. As an autarkic optimizer, I should increase my fishing time by 10%, assuming others stand pat. Now suppose, alternatively, that I think, 'If I increase my labor by 10% it's only fair that I should allow others do so; would I prefer the allocation of fish and labor that I'd enjoy under that counterfactual, where we *all* increase our labor by 10%?' Asking this question forces me to internalize the congestion externality—for although increasing my own fishing time by 10% will have only a trivial effect on the productivity of the lake, if everyone increases his labor by 10%, the congestion effect will reduce my expected take. I call this optimization protocol *Kantian*, for obvious reasons, for I take an action of the contemplated sort if and only if I would have all others take

that action as well. The corresponding stable point is a *Kantian equilibrium*: it is an allocation of fishing times to all fishers, with the property that *nobody* would advocate that *everybody* change his fishing time by *any* fraction (greater or less than one).

One can show that such an allocation of fishing times always exists, and furthermore, that it generates a *Pareto efficient allocation of fish and labor*—regardless of the particular preferences of the fishers over fish and labor, which are arbitrary, subject only to the usual premise of concavity. Therefore, if everyone employs Kantian reasoning, the tragedy of the commons is resolved! I view Kantian optimization as a micro-foundation of cooperation. Note that I have not assumed that fishers care about others—they are, in the story, only self-interested. But if the Kantian optimization protocol is in place, this cooperative way of thinking generates a Pareto efficient allocation.

Now suppose that the utility function (over fish and labor) correlates with individual fitness—it measures physical vigor, which is indeed a function of fish consumed and labor expended. Suppose there are many fishing communities, and some of them discover, perhaps because they have a clever priest or shaman, the Kantian optimization protocol, and it becomes a cultural meme, which fishers teach each other and their children. Those communities will find Pareto efficient allocations, and will be more fit than communities that are caught in the autarkic optimization protocol leading to Nash equilibrium. So we might expect to see the Kantian communities spread and the Nash communities wither, through a process of group selection.

Let's now look at hunting communities, which divide the spoils of the hunt equally among all hunters, a different allocation rule from the fishers' rule. Is there a kind of Kantian protocol for them, which induces a Pareto efficient allocation of labor and meat? Yes there is, but it is different from the one I've just described in the fishing community. The autarkic optimizing hunter says: 'I'd like to reduce my labor by two hours, and take a nap under that bush, assuming others continue hunting.' As I said, this kind of reasoning generates a Pareto inefficient Nash equilibrium for the hunting economy. But suppose the hunter thinks, 'I can only reduce my labor by two hours if I'd prefer the situation where every hunter reduces his labor by two hours. Under that assumption, would I like to do so?' The stable point associated with this optimization protocol is a vector of hunting times with the following property: *no* hunter would prefer that *all* hunters add (or subtract) *any* constant to his hunting time. I call such an allocation an *additive* Kantian equilibrium—while I now anoint the fishing equilibrium, a *multiplicative* Kantian equilibrium. One can easily show that the additive Kantian equilibrium of the hunting game is always Pareto efficient—again, regardless of the particular preferences of the individual hunters over meat and labor (see Roemer 2010; 2014).

A similar evolutionary story shows that the cultural discovery of additive Kantian optimization should spread in hunting societies. For those communities that discover it will thrive, having eliminated the inefficient equilibria generated by autarkic reasoning.

Do the members of actual fishing and hunting communities optimize in the Kantian way? If they do, then those equilibria are self-reinforcing. When Elinor Ostrom observed that many fishing communities have discovered solutions to the tragedy of the commons, she assumed that the regulations of fishing times which they enact have to be enforced with punishments and ostracism of those who disobey them. This was a natural move, because the only concept in her toolkit was Nash equilibrium. Another example of this kind of reasoning is found in Mancur Olson's famous book *The Logic of Collective Action* (Olson 1965). Olson said that cooperative behavior—that is, non-Nash behavior—could only be sustained with regulation, side payments, penalties, and punishments against deviators. But it may be the case that, in successful examples of cooperation, most or many participants are thinking in the Kantian manner, and for them, no threat of punishment or ostracism is required to enforce the equilibrium. Penalties are necessary only to keep Nash optimizers in line, of whom there may be relatively few.

In other words, I claim that autarkic (Nash) reasoning is not synonymous with rationality: it is a particular optimization protocol. I claim that Kantian optimization is no less *rational*. Rationality, recall, means a propensity to maximize a preference order. But the best way to do this, in a situation in which each individual's actions impose benefits or costs upon others, is an open question, which is not settled definitively by insisting upon the autarkic protocol. I believe Kant would agree that the kind of optimization I have named after him comprises rationality in the games in question—because it generates a Pareto efficient outcome.

These ideas, I think, are applicable to our own societies. Certainly we often teach our children to optimize in the Kantian manner: 'Don't throw your candy wrapper on the sidewalk—how would you feel if everyone did so?' Consider the payment of income taxes. I pay my taxes not because I fear the consequence of the penalty if I do not, and am caught, but because I prefer the outcome when everyone pays his taxes to the outcome when none of us do. This is true even if I am self-interested: I *need* not care about others to prefer a society in which we all pay our taxes. There are many other examples.

I propose that learning the Kantian protocol might well be part of a solution to the problem of design that concerned Jerry Cohen. Suppose that we continue to use markets—perhaps even with private ownership of productive assets. We continue to tax incomes, consumption, and wealth. But we learn to optimize in the Kantian manner. It is argued by economists, that, as an economic fact, increasing income tax rates will, at sufficiently high tax rates, reduce labor supply, as each worker chooses his labor to maximize his utility, subject to the tax rate. The additional assumption in that argument is that all workers are optimizing autarkically. When I ask myself whether I'd like to maintain my labor supply, in the face of an increase in the tax rate that has just been legislated, suppose I ask, 'Would I like to maintain my former labor supply assuming all others do likewise?' This forces me to *internalize* the effect on total tax receipts, and hence on the supply of public goods, and so on, of reducing my labor. It turns out that additive Kantian equilibrium in the tax economy induces much smaller

deadweight losses than Nash equilibrium. Although the Kantian protocol does not entirely eliminate the inefficiencies associated with income taxation, it reduces them substantially. This is what I was referring to when I said earlier that we might be able to do *much better* than the Nordic economies are currently doing, assuming that many, in those economies, still optimize in the autarkic manner when determining their occupational choices or labor supplies.

As I said, the Kantian optimization protocol can be viewed as providing micro-foundations for cooperative behavior. It is a micro-foundation in the sense that it is a theory of individual action that leads to cooperation, to coordinating economic activity to eliminate the waste inherent in Pareto inefficiency that plagues Nash equilibria.

What is the relationship between Kantian optimization and socialist ethos? This is an open question. As I mentioned, I believe that the next step on the way to the spread of a socialist ethos is the achievement of quantitatively more cooperation than currently characterizes our societies. My conjecture, or perhaps it is only a hope, is that that achievement will generate the spread of socialist ethos. First of all, increased cooperation will hasten our escape from the period of material privation, and as I've said, I think that that escape will potentiate the evolution of socialist ethos. Secondly, I think that in achieving cooperation, we will see that we all have better lives: and showing that *possibility* will facilitate the adoption of socialist ethos. I suppose I believe in a kind of 'ought implies can' law of motion for social ethos: showing that we *can* improve all our lives if we cooperate means that we *ought* to care about the lives of others as well as our own. More precisely, I should say that a necessary condition for our holding the socialist ethos is the demonstration that we *can* massively improve all our lives, together, if we cooperate. My conjecture is that understanding this, plus the disappearance of material scarcity, will be jointly sufficient. But I have no proof... yet.

The focus in this view upon overcoming material deprivation requires me to mention another theme of Jerry's: and that is his comments upon the relationship between communism and material abundance. Jerry wrote that Marxists have heretofore believed that communism would occur when technology became sufficiently productive to overcome material scarcity, but that, given the constraints imposed upon us by limited natural resources, we must revise that rosy view. Climate change associated with atmospheric carbon concentration is the most dramatic example of the natural constraint that we face today. We must ask whether natural constraints will prevent our overcoming the period of material scarcity in human history, which I have suggested is a pre-condition to the universal development of socialist ethos.

Economists tend to be optimists: we understand the power of compound interest, and in particular, how technological change can, over the course of a few centuries, dramatically alter the human condition. I share this optimism. I do not, however, believe the climate-change problem is a trivial challenge: I think it is the most severe problem that mankind faces at present. The reason, however, is not technological or natural, but political. From work that I have been doing in recent years with a small team of collaborators, I believe that

it is technologically feasible to continue to improve the human condition, while not generating catastrophic temperature change (see Llavador/Roemer/Silvestre 2015).

There are two major political problems. The first is overcoming the power of the energy industry, whose leaders have no desire to eliminate fossil fuels, and the second is finding an international cooperative solution that will permit the global South to develop rapidly, while not destroying the biosphere. To reiterate, I think both these problems are technologically and physically solvable: that is, there exist feasible paths of resource use, that will permit the global South to converge to the global North in per capita welfare, will not destroy the biosphere, and will, furthermore, not require *reductions* in the living standards of those in the global North. But implementing one of these feasible paths *will* require state intervention in the economy in ways that are anathema to the US Republican Party as it is currently constituted, *will* require transition away from fossil fuels, and *will* require reducing the rate of growth of living standards in the North below what its citizens have come to expect. Indeed, political feasibility may well require substantial redistribution of income from the rich to the rest in the North, to gain citizens' assent to lower growth. Whether we can achieve these tasks is, in my opinion, the big question for social science in the coming period, but it is, luckily, a question which I need not address here.

## Bibliography

- Berman, S. (2011), Social Democracy and the Public Interest, in: *Critical Review* 23, 237–256
- Bowles, S./H. Gintis (2011), *A Cooperative Species: Human Reciprocity and its Evolution*, Princeton
- Cohen, G. A. (1986), Self-ownership, World-ownership and Equality, in: Lucash, F. S./J. N. Shklar (eds.), *Justice and Equality Here and Now*, Ithaca
- (1989), On the Currency of Egalitarian Justice, in: *Ethics* 99, 906–944
- (2008), *Rescuing Justice and Equality*, Cambridge/MA
- (2009), *Why not Socialism?*, Princeton
- Llavador, H./J. E. Roemer/J. Silvestre (2015), *Sustainability for a Warming Planet*, Cambridge/MA
- Nozick, R. (1974), *Anarchy, State and Utopia*, New York–Oxford
- Olson, M. (1965), *The Logic of Collective Action*, Cambridge/MA
- Ostrom, E. (1990), *Governing the Commons*, New York
- Pinker, S. (2011), *The Better Angels of our Nature*, New York
- Rawls, J. (1971), *A Theory of Justice*, Cambridge/MA
- Roemer, J. (2010), Kantian Equilibrium, in: *Scandinavian Journal of Economics* 112, 1–24
- (2014), Kantian Optimization: A Micro Foundation for Cooperation, in: *Journal of Public Economics*, URL: <http://dx.doi.org/10.1016/j.jpubeco.2014.03.011>