Do Kantians drive others to extinction?

Jean-François Laslier

PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

WORKING PAPER N° 2020 – 28

Do Kantians drive others to extinction?

Jean-François Laslier

JEL Codes:
Keywords:

Funded by a French government subsidy managed by the ANR under the framework of the Investissements d'avenir programme reference ANR-17-EURE-001

Do Kantians drive others to extinction?

Jean-François Laslier
CNRS, Paris School of Economics*

May 22, 2020

Abstract

I comment on the claim by John Roemer that "in games of pure coordination, Kantians drive Nashers to extinction". Using an explicit dynamic model of evolution, I notice that in these games, Kantian optimizers do not always drive selfish optimizers to extinction.

1 Introduction

From Malthus and Verhuulst to Volterra and Lokta, theoretical biology has studied the dynamics of living species (see Berrymann, 1992, for an account of this history). Its interaction with game theory (Maynard Smith, 1982) has been fruitful, and a whole discipline of evolutionary game theory has emerged (Weibull, 1995).

Of particular interest is the explanation of cooperative behavior and altruism based on evolutionary arguments, because the existence of cooperation may seem at first sight in contradiction with the individual selection paradigm in biology. But, as Roemer recalls in his book, men (and animals too) routinely behave in a cooperative manner, sometimes even at their own expense. This explains why the question of cooperation has been a non-trivial puzzle for evolutionary biology (see the work of Hamilton, 1963, 1964, and more recently Nowak and Sigmund, 2005 or Alger and Weibull, 2013). It is also a central question for the economic theory, all the most since the question of incentives and selfish behavior became prevalent in mainstream economics. In the eighth chapter of his book *How We Cooperate*, Roemer applies the concept of Kantian optimization to coordination games in order to offer an evolutionary view of this concept, compared with what he calls the "Nash behavior".

Coordination requires giving some attention to what the others do, and this is of course one element of cooperation. Games where individuals may settle on

*École Normale Supérieure, 48 Bd. Jourdan, 75014 Paris, France. jean-francois.laslier@ens.fr. I wish to warmly thank Vaios Koliofotis and Marina Uzunova who spotted and fixed and important mistake in the first version of this paper. All remaining errors are mine.

a low quality outcome while coordination on a better one is also possible are therefore interesting school-cases for the study of cooperative behavior, even if cooperation should not be reduced to efficient coordination. The point was introduced by Rousseau (1755) in his *Discours sur l'origine et les fondements de l'inégalité parmi les hommes*. The frame is an evolutionary perspective, but a pre-Darwinian, non modern, one. The well-known "stag-hunt" game is presented by Rousseau as an illustration of the fact that behavioral coordination requires some kind of language, but a language that can be restricted to specific signaling goals, as animals have.

Roemer's analysis of the stag hunt game incorporates one important idea of modern evolutionary theory, namely the concept of evolutionary stable equilibrium (EES). It also alludes to the possible dynamics that sustain the "stability" of these equilibria. There are no explicit dynamics in Roemer's construction, but, as the name tells, they are a key feature of evolutionary theories. These dynamics make no reference to Rousseau's idea of communication as a mean for coordination but instead model some Darwinian process of selection of the fittest.

In the exercise performed in chapter eight of *How We Cooperate*, the non-Kantian agents are called "Nashers". The word refers to an equilibrium concept, the "Nash equilibrium," that, unlike the EES, has no associated dynamic process that would insure some form of stability. The cocktail proposed in chapter eight, that compares the fate of Nashers and Kantians along their evolutionary dynamics therefore needs to be spelled out more precisely. Additional insights can be provided by a double dynamics model that takes into account both the dynamics of selfish optimisation, which might sustain Nash behavior, and the dynamics of selection or survival of Kantians. This what I do in the following lines.

2 Roemer's model

Althought Roemer discusses symmetric coordination games in some generality, I will concentrate on one example: the so-called stag hunt game.

2.1 The stag hunt game

Rousseau tells a story about how men were gradually able to acquire the concept of mutual commitment.

> In this manner, men may have insensibly acquired some gross ideas of mutual undertakings, and of the advantages of fulfilling them: that is, just so far as their present and apparent interest was concerned: for they were perfect strangers to foresight, and were so far from troubling themselves about the distant future, that they hardly thought of the morrow. If a deer was to be taken, every one saw that, in order to succeed, he must abide faithfully by his post: but if a hare happened to come within the reach of any one of them, it is not

> to be doubted that he pursued it without scruple, and, having seized his prey, cared very little, if by so doing he caused his companions to miss theirs.
> It is easy to understand that such intercourse would not require a language much more refined than that of rooks or monkeys, who associate together for much the same purpose.[1]

A standard version of this game is the one described by Roemer, on pages 29 and 128 of his book. There are two hunters who belong to the same population. If they hunt separately they will grab hares, and this is worth the normalized payoff 0. If they together hunt the stag they will each earn the best payoff, say 1. Now if only one hunts the stag, he will catch neither stag nor hare, and will therefore earn a negative payoff, say -1; meanwhile the other hunter chasing hares will obtain more hares than if both where hunting hares, and this is valued .5. That is Roemer's (a, b) game for $a = -1$ and $b = .5$. This game is a classic of game theory; for instance Skyrms (2004) sees it as a fable that describes the key feature of the social contract, and Ken Binmore uses it to defend his claim that "fairness evolved as Nature's answer to the equilibrium selection problem in the human game of life" (Binmore 2006).

Table 1: The game in normal form

	stag	hare
stag	$(1, 1)$	$(-1, .5)$
hare	$(.5, -1)$	$(0, 0)$

2.2 Kantians

It is clear that hunting the stag is the thing to do for efficiency reasons. The unique Kantian equilibrium in this game is hunting the stag because it is better for everyone if everyone does the same. The key concept of Roemer's analysis is very clear in this game: any Kantian player hunts the stag.

2.3 Nashers

This game has two pure strategy equilibria:

[1] Translation by G. D. H. Cole. See http://www.constitution.org/jjr/ineq.htm. The exact quote is: Voilà comment les hommes purent insensiblement acquérir quelque idée grossière des engagements mutuels, et de l'avantage de les remplir, mais seulement autant que pouvait l'exiger l'intérêt présent et sensible; car la prévoyance n'était rien pour eux, et loin de s'occuper d'un avenir éloigné, ils ne songeaient pas même au lendemain. S'agissait-il de prendre un cerf, chacun sentait bien qu'il devait pour cela garder fidèlement son poste; mais si un lièvre venait à passer à la portée de l'un d'eux, il ne faut pas douter qu'il ne le poursuivît sans scrupule, et qu'ayant atteint sa proie il ne se souciât fort peu de faire manquer la leur à ses compagnons. Il est aisé de comprendre qu'un pareil commerce n'exigeait pas un langage beaucoup plus raffiné que celui des corneilles ou des singes, qui s'attroupent à peu près de même.

- Both players hunting the stag is a strict Nash equilibrium because it pays 1 and looking alone for hares pays only .5.
- Both players hunting hares is a strict Nash equilibrium because it pays 0 and looking alone for a stag pays -1.

Following Harsanyi and Selten (1988), hunting the stag is called the *payoff dominant* equilibrium and hunting hares the *risk dominant* equilibrium. The game has also a mixed-strategies Nash equilibrium:

- Both players deciding at random independently to hunt the stag with probability 2/3 and hares with probability 1/3 is a Nash equilibrium because if a player uses this mixed strategy, then the average payoff for the other player is the same whatever he does. (The second player thus has no strict incentive to choose one strategy rather than the other and so, under the usual hypothesis about choice under uncertainty, he can as well himself randomize the same way.)

The definition by Roemer of a “Nasher” might not be perfectly clear in general. The word is used in place of the expression “Nash optimizer” (p. 117) but the meaning of what is a “Nash optimizer” is only clear with respect to a given Nash equilibrium, not with respect to the game itself nor with respect to the other players’ strategies. Roemer writes (page 118) that “If there are several Nash equilibria, a Nasher randomizes among them”. This is difficult to follow: it is unclear whether this means that different Nashers end up playing different strategies, or that they manage to correlate their randomization in order to all play the same (randomly chosen) pure strategy. Moreover, in some instances, as in the game above, each one of the two pure strategies is played in a Nash equilibrium, but choosing at random does not, in general, result in an equilibrium. In fact, except for a very specific randomization scheme, mixed strategies are almost never best responses, and are therefore almost never chosen by an optimizer.

Therefore, in order to understand what can a “Nasher” be, one has to see how this concept is used.

2.4 Roemer’s evolutionary argument

The argument that leads to the conclusion that “Kantians drive Nashers to extinction” is provided in the proof of Proposition 8.4 (page 123). The proof first fixes the Nash equilibrium to be considered and denotes the associated strategy by q^*. (Thus the Nashers do not randomize among different equilibria.) A Nasher hunts the stag with probability q^* and the hares with the complement probability $1 - q^*$. Two cases are considered: $q^* = 0$ (hunting hares) and the mixed equilibium $q^* = q_1^*$ that has, in this game, the player hunting the stag with probability $q_1^* = 2/3$ and hares with probability 1/3, yielding the average payoff 1/3 in any case. According to Roemer, the third case (Nashers hunting the stag) is not to be considered because in that case Nashers and Kantians cannot be distinguished.

Following the standard evolutionary model, one imagines that individuals are randomly matched by pairs (with no "assortative" matching). Let ν be the proportion of Kantians in the whole population, then the average payoffs for a Kantian ($V^K(\nu)$) and for a Nasher ($V^N(\nu)$) can be computed in the two cases:

- If Nashers hunt hares ($q^* = 0$) :

$$\begin{array}{lllllll} V^K(\nu) & = & \nu \cdot 1 & + & (1-\nu)\cdot(-1) & = & 2\nu - 1 \\ V^N(\nu) & = & \nu \cdot \frac{1}{2} & + & (1-\nu)\cdot 0 & = & \nu/2. \end{array}$$

- If Nashers use the mixed strategy $q^* = q_1^* = 2/3$:

$$\begin{array}{lllllll} V^K(\nu) & = & \nu \cdot 1 & + & (1-\nu)\cdot \frac{-1}{3} & = & (2+4\nu)/6 \\ V^N(\nu) & = & \nu \cdot (\frac{2}{3} + \frac{1}{3}\cdot\frac{1}{2}) & + & (1-\nu)\cdot \frac{1}{3} & = & (2+3\nu)/6. \end{array}$$

Now, it is clear that in the case where Nashers use the mixed strategy q_1^*, Kantians have an evolutionary advantage, because their payoff is larger than the payoff to Nashers: $(2+4\nu)/6 \geq (2+3\nu)/6$.

If Nashers hunt hares, then the advantage will be on one side or the other depending on the value of ν: Kantians have the advantage if and only if ν is large enough. In words: if Nashers hunt hares but most players hunt the stag, Nashers earn less on average than the others.

This is Roemer's argument to claim that Kantians drive Nashers to extinction.

3 A double dynamics model

In the above argument, the "evolutionary analysis" is only sketched. In particular the "Nashers" are neither optimizing agents (as they should be in an economic model of rational behavior) nor adapting (as they should be a behavioral model of learning) nor evolving (as they should be in a biological model of darwinian selection), they are just stubborn hare-hunters in one case and (quite strangely) stubborn users of a specific mixed strategy in the other case. I now propose a standard model of the replicator dynamics type, in order to study, for this game, the evolution of a population made of Kantians individuals (in Roemer's sense) and of adaptive ones. I will simple call "selfish" the non-Kantians, although one could think of many names for them.

3.1 Replicator dynamics

What follows uses the most standard mathematical model of evolution called the "replicator" dynamics. So I begin by a very brief presentation of this model.

The *fittness* defines the number of offspring of the some replicating unit as a function of its environment, so that a population of size n characterized by a fitness per individual f grows at the rate f. In discrete time the population of size n will be of size $f \cdot n$ at the next generation, and in continuous time, the time derivative of n is $\dot{n} = \mathrm{d}n/\mathrm{d}t = f \cdot n$ with f being now a replication rate by unit of time.

With two groups $i = 1, 2$ of size n_i and fitness f_i, writing $n = n_1 + n_2$ and $x_i = n_i/n$ one gets, in full generality:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{n_i}{n}\right) = \frac{\dot{n_i}n_j - \dot{n_j}n_i}{n^2}$$

that is:

$$\dot{x_i} = x_i x_j \cdot (f_i - f_j).$$

Interestingly, this differential equation , that generates "replicator dynamics" appears alike in different models that describe (i) population genetics, (ii) social imitation, or (iii) individual adaptive learning (see Laslier et al. 2006). We will now apply this idea to obtain an evolutionary model for Roemer's argument.[2]

In the absence of Kantians, the standard evolutionary analysis of the stag-hunt games indicates that a pure strategy equilibrium is reached in the long run: if the initial composition of the population contains enough individuals of one type (be it stag hunters or hare hunters) coordination on this type will occur in the long run. The mixed strategy equilibium is, on the contrary, unstable and is not reached. To introduce the Kantian players, I propose the following model:

Let $\nu(t)$ be the proportion of Kantians in the population at time t. This proportion will vary with time. Following the evolutionary paradigm, the non-Kantians will not be supposed to jump directly to some optimal or "Nash" strategy, but they will adjust their strategies gradually with time. So let $x(t)$ denote the proportion of stag-hunters among the non-Kantians. In the whole population the proportion of stag-hunters is therefore:

$$y = \nu + (1 - \nu)x.$$

Two processes of evolution are coupled: within the non-Kantians for their choice of strategy, and between Kantians and non-Kantians. The two processes may occur at different speed.[3] For instance one may wish to study the case where selfish individuals can change strategy relatively quickly while it is only at a slow pace that selfish individuals become Kantians or Kantians become selfish. This is rather natural: it means that selfish individuals adjust behavior by choosing a best response to the circumstances if not instantly, at least relatively quickly. The definition of a Nasher by Roemer does not presupposes a dynamic adjustment process, the underlying assumption is that Nashers instantly find best responses. Therefore in order to relax this assumption we will consider

[2]Note that an important literature exists, in evolutionary game theory, about the stag hunt problem and extension of the basic game (Kandori et al. 1993; Samuelson 1997). The main focus is on the question of communication. Does language help to coordinate on the payoff dominant equilibrium ? Can one even explain the emergence of language as a coordinating device that allows forward induction in stag-hunt situations? See Chapter 8 in Samuelson (1997).

[3]The same idea of two-level dynamic process is used in Laslier and Göktuna (2016) and in Göktuna (2019).

that this process is not instantaneous but goes relatively fast compared with the transition from selfish to non-selfish. In our model, individuals can be viewed from the perspective of two-time scales. In the long-run or evolutionary time, it takes many generations to converge on a long-run equilibrium in accord with the dynamic process of natural selection (a slow process of change). Being a Kantian or a selfish optimizer is determined by this long term evolution. In the short run or decision-making time, rational individuals make choices or acquire new strategies via social learning (a fast process of change). The choice of her strategy by a selfish optimizer takes place in this shorter term. Hence, the overall evolutionary process has the complex structure of a slow evolutionary and a fast decision-making time horizon.

Among selfish individuals, the difference in payoff between those who hunt the stag and those who hunt hares is:

$$\delta^1 = [y - (1-y)] - [y/2] = 3y/2 - 1.$$

Let s be the adaptive speed of the selfish individuals. The replicator dynamics within this group is described by the following differential equation (where $\dot{x}$ denotes the time derivative $\mathrm{d}x/\mathrm{d}t$):

$$\dot{x} = s \cdot x(1-x) \cdot \delta^1 \tag{1}$$

At the level of the whole population, the difference in payoff between Kantians and selfish individuals is obtained as follows. For the Kantians, since they all play the same strategy (stag), the average payoff is simply the average payoff of the stag strategy, that is $2y - 1$. For the other group, one should think of them as carriers of a "selfish" gene, whose fitness is the average fitness of the individuals who carry it; therefore the payoff to take into account for the evolution of the selfish population is the average payoff in this population, that is:

$$x \cdot (2y-1) + (1-x) \cdot y/2.$$

hence

$$\delta^2 = [2y-1] - [x(2y-1) + (1-x)y/2] = (1-x)(3y/2-1).$$

We can normalize at 1 the adaptive speed of kantianism, then the associated replicator dynamics is described by the following differential equation :

$$\dot{\nu} = \nu(1-\nu) \cdot \delta^2. \tag{2}$$

The quantities y, δ^1 and δ^2 depend solely on the variables ν and x, so the equations (1) and (2) define a system of differential equations in the square $(\nu, x) \in [0,1] \times [0,1]$.

3.2 Results

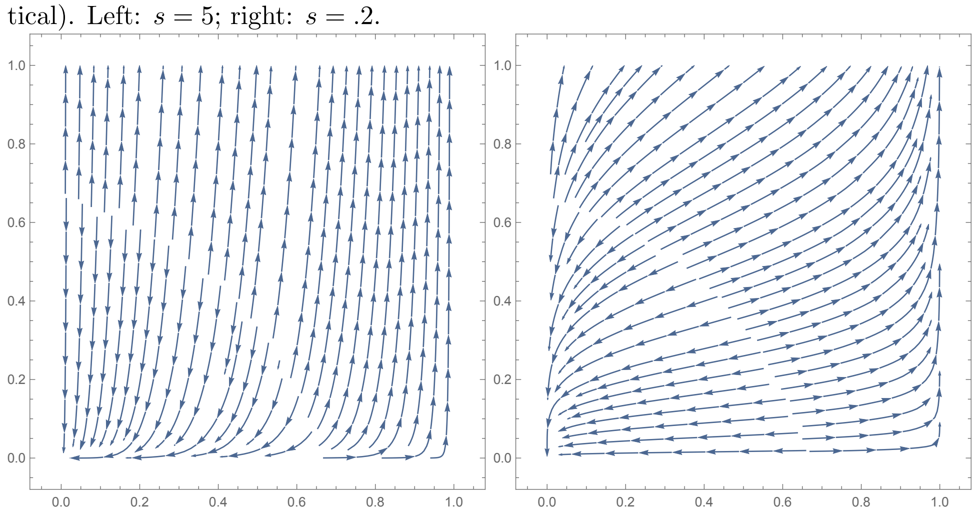
All the results can be read on Figure 1. The left panel of the figure is drawn for a speed ($s = 5$) such that selfish individuals adapt their strategy relatively quickly. This is the most natural assumption. On the horizontal axis is the proportion ν of Kantians and on the vertical axis is the proportion x of stag-hunters among the population of selfish individuals. The picture represents the flow of the differential system.

The lower left corner $(0, 0)$ corresponds to the situation where there are non Kantians and everyone is hunting hares. The upper right corner corresponds to the situation where all the population is made of Kantians and everyone hunts stags. The upper segment, where $x = 1$, also describes situations of full cooperation, where everyone hunts stags but some do it because they are Kantians and others do it for selfish reasons.

Following the arrows, one can see the fate of the system. The flow is divided in two: a lower left region that points to $(0, 0)$ and an upper right region that always reaches the upper segment. Changing the speed, as it is done in the right panel of Figure 1, confirms this point.

The two basins of attraction are separated by the curve of equation $\dot{\nu} = 0$, that is $3y/2 = 1$ or, in terms of ν and x: $3\nu + 3(1 - \nu)x = 2$. This is part of the hyperbola $x = \frac{2-3\nu}{1-\nu}$ and is independent of s.

Figure 1: Stag-hunt game: coupled dynamics for the proportion of Kantians (horizontal axis) and the proportion of cooperators among non-Kantians (vertical). Left: $s = 5$; right: $s = .2$.



4 Conclusion

In section 8.2. of his book, after the study of the stag hunt game in isolation, Roemer considers several games. He obtains that if Nature chooses at random what kind of game is played, either a coordination game or a prisoner's dilemma, then Nashers and Kantians can co-exist. This note focused instead on the single coordination game.

Reading the claim (p.125) that "In games of pure cooperation, Kantians drive Nashers to extinction", readers of Roemer's book might be tempted to over-interpret the expression "Nasher" and to read that Kantian optimizers (in the sense of Laffont and Roemer) have some efficiency advantage, in coordination games, that make them better fitted, from the evolutionary point of view, than selfish optimizers. This is not true. In the stag-hunt game either the Kantians are wiped away by selfish individuals who do not cooperate, and the Kantians are "driven to extinction" by the selfish optimizers, or both remain as some fraction of the population.

Focusing exclusively on "equilibria" to describe and analyze collective outcomes may be misleading. Following his analysis, Roemer writes (pages 123) that "According to Proposition 8.4. there are no (a, b) games where both Kantian and Nash players exist with positive frequencies in a stable equilibrium." It is not clear what is meant there by "stable equilibrium" but, as I showed above, the natural process that sustains evolutionary stability leads to two possible outcomes, depending on the initial conditions. One possibility is that hare hunters (who can be called Nashers) drive Kantians stag hunters to extinction. The other possibility is that Kantians and selfish optimizers (who aslo can be called Nashers) co-exit, all hunting stags but for different reasons.

References

[1] Ingela Alger and Jörgen Weibull (2013) "Homo Moralis. Preference evolution under incomplete information and assortative matching" *Econometrica* 81: 2269—2302.

[2] Alan A. Berryman (1992) "The origins and evolution of predator-prey theory" *Ecology* 73(5): 1530–1535. doi:10.2307/1940005.

[3] Ken Binmore (2006) "The Origins of Fair Play" *Papers on Economics and Evolution* No. 0614, Max Planck Institute of Economics, Jena.

[4] Jean-Jacques Laffont (1975) "Macroeconomic constraints, economic efficiency, and ethics: An introduction to Kantian economics" *Economica* 42: 430—437.

[5] William D. Hamilton (1963) "The Evolution of Altruistic Behavior." *The American Naturalist* 97 (896): 354–356.

[6] William D. Hamilton (1964) "The Genetical Evolution of Social Behaviour I and II." *Journal of Theoretical Biology* 7 (1): 1–52.

[7] John C. Harsanyi and Reinhard Selten (1988) *A General Theory of Equilibrium Selection in Games*, MIT Press.

[8] Michihiro Kandori, George Mailath, and Rafael Rob (1993) "Learning, mutation, and long-run equilibria in games" *Econometrica* 61: 29—56.

[9] Jean-François Laslier, and Bilge Öztürk-Göktuna (2016) “Opportunist politicians and the evolution of electoral competition” *Journal of Evolutionary Economics* 26(2): 381–406.

[10] Jean-François Laslier, Gisèle Umbhauer and Bernard Walliser (2006) “Game situations.” Chapter 3, pages 67–112 in: *Evolutionary Microeconomics*, edited by J. Lesourne, A. Orléan and B. Walliser. Springer.

[11] John Maynard Smith (1982) *Evolution and the Theory of Games*. Cambridge Universiy Press.

[12] Martin A. Nowak and Karl Sigmund (2005) “Evolution of indirect reciprocity” *Nature* 437: 1293—1295.

[13] Bilge Öztürk-Göktuna (2019) “A dynamic model of party membership and ideologies” *Journal of Theoretical Politics* 31 (2): 209—243.

[14] John Roemer (2019) *How We Cooperate*. Yale University Press.

[15] Jean-Jacques Rousseau (1755) *Discours sur l'origine et les fondements de l'inégalité parmi les hommes*. Reprinted in : *Ecrits politiques*, 1992, Le Livre de Poche.

[16] Larry Samuelson (1997) *Evolutionary Games and Equilibrium Selection*. MIT Press.

[17] Brian Skyrms (2004) *The Stag Hunt and the Evolution of Social Structure*. Cambridge University Press.

[18] Jörgen Weibull (1995) *Evolutionary Game Theory*. MIT Press.