

Capitalism According to Evolutionary Game Theory: The Impossibility of a Sufficiently Evolutionary Model of Historical Change

YANIS VAROUFAKIS*

ABSTRACT: Evolutionary game theory has recently furnished some exciting theoretical and experimental insights regarding the birth of social power and discrimination. But can this type of theory illuminate the history and nature of capitalism? The answer turns out to be negative: evolutionary models are bound to remain either insufficiently evolutionary or hopelessly indeterminate. However, social theorists have much to gain from understanding what would breathe social life into evolutionary game theory's models: a proper historical account of the sources of behavioral variety, and an adaptive mechanism that leaves room for the cunning of human reason.

1. *The Case for Taking Evolutionary Game Theory Seriously*

IN A FAMOUS LETTER TO ENGELS, Marx commented that it was "remarkable how Darwin rediscovers, among the beasts and plants, the society of England with its division of labor, competi-

* Thanks are due to Euclid Tsakalotos, Geoff Hodgson and Tony Aspromourgos for helpful comments on an earlier draft of a related paper presented at the two-day conference on "The Variety of Economic Institutions under the Many Forms of Capitalism," Athens University of Economics and Business, May 12–13, 2006. I am also grateful to Ioanna Minoglou for giving me the impetus for this paper as well as to participants in seminar presentations at Birkbeck College, London and at the Catholic University of Louvain-la-neuve. All errors and incongruities are mine.

tion, opening up of new markets, 'inventions' and Malthusian 'struggle for existence'.¹ Nevertheless, it is quite telling that Engels, despite his deep awareness of Marx's criticism of Darwin, should choose to farewell his great friend, on that bleak afternoon in London, with the now infamous words: "Just as Darwin discovered the law of development of organic nature, so Marx discovered the law of development of human history."²

Thus a delicate, powerful and edgy relationship began between two of the most significant intellectual currents to have emerged out of the late 19th century: Darwin's evolutionary theory and Marx's historical materialism. For decades the two danced around each other, sometimes in harmony, at other times in brutal opposition. The 20th century's cruelty played a definite part in keeping the two at an arm's length,³ while the 21st confuses the relationship further as evolutionism's politics differ depending on the field in which it is applied. Thus, while in biology it is engaged in the good struggle against assorted forces of darkness (*e.g.*, Creationism, "Intelligent Design," etc.), in the social sciences it is often to be found in an alliance with the

- 1 Marx to Engels, Manchester, March 29, 1865 (Marx and Engels, 1979, Vol. 41, 380). In his introduction to *Origin of the Species* (Darwin, 1859), Darwin acknowledged the influence on his thinking of classical political economy by referring explicitly to Malthus' apocalyptic theory of population pressure on the means of subsistence. Indeed, Darwin introduces his readers to his now legendary theory of the struggle for existence as no more than an extension of Malthus' economic theory to "the whole animal and vegetable kingdoms" (4-5). Malthus was concerned that human population grew geometrically while food production could only grow arithmetically. If so, a struggle for existence would occur as increasing numbers of people would have to starve. Darwin was clearly impressed by this. In his own words: "In the next chapter the Struggle for Existence amongst all organic beings throughout the world, which inevitably follows from the high geometrical ratio of their increase, will be treated of. This is the doctrine of Malthus applied to the whole animal and vegetable kingdoms" (1860, 4-5). Interestingly, in a much acclaimed recent book, Vermeij (2006) continues Darwin's legacy by "discovering" that Nature is evolving in a manner uniquely captured by mainstream economics!
- 2 Engels' complete graveside reference to the parallelism of Marx and Darwin follows: "Just as Darwin discovered the law of development of organic nature, so Marx discovered the law of development of human history: the simple fact, hitherto concealed by an overgrowth of ideology, that mankind must first of all eat, drink, have shelter and clothing, before it can pursue politics, science, art, religion, etc.; that therefore the production of the immediate material means, and consequently the degree of economic development attained by a given people or during a given epoch, form the foundation upon which the state institutions, the legal conceptions, art, and even the ideas on religion, of the people concerned have been evolved, and in the light of which they must, therefore, be explained, instead of vice versa, as had hitherto been the case" (Engels, "Speech at the Graveside of Karl Marx," 1883).
- 3 The use to which the Nazis put Darwinian logic was one reason for the left's reaction. The Soviet approach was, however, also indefensible. Pollock (2007; see ch. 6) gives a flavor of Stalin's attitude toward biologists.

worst variants of a conservatism whose *raison d'être* is the legitimization of a Panglossian view of late capitalism.

This paper argues that progressive social theorists, especially those who are sceptical towards social evolutionism, need to take a close look at its latest incarnation: *Evolutionary Game Theory* (EvGT hereafter). Before offering four reasons for this, I shall first dispel two frequently aired objections to the application of the logic of evolution to social theory.

First, there is the objection that to entertain the logic of evolution within the social sciences is to impose biological applications on the canvas of human history. Second, that a theory of automated responses cannot be but misleading, if not tautological, when applied to rational agents. Neither objection holds water.

The first confuses the historical origins of evolutionary theory with its essence.⁴ An evolutionary argument is based on the idea that systemic change occurs through two related processes: 1) the causal mechanism that generates a *variety* of behaviors or types within a class of entities; and 2) a dynamic process of *selection* which weeds out most of those behaviors and types, therefore yielding ever-changing, yet clearly discernible, patterns. It is quite obvious that nothing in this logic, of the incessant march of variety and selection, is tied specifically to biology. That political economy shied away from evolutionary reasoning, aligning itself to a type of crude classical mechanics, and leaving the field of evolutionary theorizing to biology, is surely to be blamed neither on evolutionary thinking nor on biology.⁵

The second objection, related to the possibility that the miracle of human reason cancels the logic of evolution, is a tougher proposition. Marx habitually, and rightly, poured scorn on those (*e.g.*, Spinoza and Feuerbach) who transplanted models from the natural to the social sciences with little or no modification to allow for the fact that human beings are ontologically different to atoms, molecules, genes

4 Indeed, Darwin himself admits to having borrowed evolutionary thinking from political economy (see note 1).

5 In fact, it is possible to argue convincingly that social theory is a more fertile ground on which to employ evolutionary logic than biology ever was! Recall that social theorists and historians do little else than to focus upon variety and selection processes (while, perhaps, calling them something different) for the purpose of explaining how: a) new social entities are born, and b) their relative importance in explaining changes over time through the elimination of some and the reinforcement of others. Meanwhile, rates of social evolution are certainly faster than rates of natural selection and, in this sense, the logic of evolution is plainer to observe in society than it is in biology.

and planets: "What distinguishes the worst of architects from the best of bees," he wrote in *Capital I*, is that "the architect raises his structure in imagination before he erects it in reality."

Granted that our theories should take seriously the fact that humans have a unique capacity to create in their heads not only buildings but also models of their own evolution, we should not forget the greatest legacy that Marx bequeathed us: The concept that our ideas do *not* shape the world, but reflect the systemic evolution of the social relations of our material reproduction; the thought that rational persons make history, but not under conditions of their choosing.⁶

The question therefore is: Can we, *both* as rational agents of change *and* as playthings of social forces beyond our control, make sense of our own history *without utilizing the logic of evolving systems*? The short answer is negative. The longer one is that, while the logic of evolution is indispensable, it is also insufficient; that evolutionary *models* are good at explaining variational change (*i.e.*, selection within a paradigm, *e.g.*, the rise of consumerism after World War II) but find it hard to inform transformational change (*i.e.*, the emergence of a new paradigm such as, *e.g.*, capitalism).

In summary, the problem with evolutionary social theory is neither that its origins are biological nor that its predictions follow trivially from its assumptions. It is rather that, in accordance with neoliberal thinking (of which mainstream economics is the best example), it tends to stay firmly within the confines of a *canonical* (that is, primitive) *evolutionary model* in which the individual is frozen in time and change occurs only at the level of the social. The study of history requires a deeper analysis, one that accounts for the *sources* of variety and connects them to the cunning of human reason.

Does this mean that existing evolutionary social theory is not worth a penetrating look? On the contrary, getting to know what would breathe social life into an evolutionary theory is the key to good practice in the social sciences. And just as the study of the limitations of Adam Smith's and David Ricardo's theories gave Marx the necessary framework in which to develop his political economy, so today the study of the limitations of the highest form of evolutionary social

6 The materialist idea that ideas are shaped by the joint evolution of means and relations of production needs no citation, as Marx's writings are replete with it. The thought that we are authors of our own life, but not under constraints of our choosing comes from *The Eighteenth Brumaire of Louis Bonaparte*.

theory, EvGT, is supremely useful to anyone who needs new arguments and fresh inspiration for a revitalization of historical materialism.

Beyond the grander reason offered in the previous paragraph, the study of EvGT is rewarding because

- (1a) EvGT is a source of much pride in the corridors of the great social science departments of our time and, moreover, because big claims are made on its behalf, not only by neoliberals, but, also, by progressive theorists who invest it with hopes of re-founding social theory on a scientific basis.⁷
- (1b) EvGT generates some interesting theoretical results which often subvert the intentions and the ideology of its assumptions (see the next section and the discussion of discrimination).
- (1c) EvGT holds useful lessons for Marxists and progressive social theorists *viz.* the prerequisites of a non-mechanistic theory of capitalist development.⁸

The next section relates EvGT's main theoretical and experimental insights. In Sections 3 and 4 I try to imagine EvGT's best attempt at narrating the evolution of capitalism. Sections 5 and 6 draw on the theoretical impasse of EvGT in order to highlight the difference between evolution and history, a difference that holds the key to the possibility of human values that are both enabled and threatened by capitalist dynamics.

2. *Evolutionary Game Theory: Its Main Insights and Some Intriguing Experimental Evidence*

EvGT evolved from the models of evolutionary biologists Maynard-Smith and Price (1974) and Dawkins (1976), which inspired game

⁷ Elster (1982, 1986), for example, has criticised Marx's utilization of functionalist argument (*e.g.*, in his theory of the state as an expression of capital's will) because they do not fill in how the unintended consequences of the individuals' actions help promote the activity which is responsible for this set of unintended consequences. His point is that there has to be a *feedback mechanism*: that is, something akin to the principle of natural selection in biology which is capable of explaining behaviors by their "success" and not by their "intentions." EvGT is often mentioned, *e.g.*, by young Marxist scholars like Roberto Veneziani, as one potential source of such a mechanism that may give Marxism a new analytical legitimacy.

⁸ I mention this because, while EvGT can be easily criticized for the simple Darwinian mechanism at its heart, Marx himself has been accused of mechanism and, indeed, in the modern (primarily Anglo-Saxon) social theory literature he is taken to be an exemplar of 19th century mechanism. The fact that he would deny this, pointing to the dialectical method he borrowed from Hegel, does not make it less important for us today to point out precisely where in the analysis the dialectic comes in. The critique of EvGT makes this task simpler.

theorists for reasons that we need not delve into here.⁹ Our interest lies in the capacity of these models to tell stories that resonate helpfully with some of the enduring puzzles of social theory. Take the case of *discrimination*. First-wave feminists, to mention one example, faced the enormous task of explaining how the highly asymmetrical social roles of men and women need not reflect biological differences. EvGT would have been their ally!

EvGT shows that, even in populations of utterly identical agents, sustainable discrimination is not only a possible result of behavioral evolution but, remarkably, that it is an *inescapable* “evolutionary equilibrium.” To make this point clearly, consider EvGT’s favorite game: the so-called *hawk-dove* game. Two players have to choose between aggressive (*h*) and cautious (*d*) behavior. Mutual aggression (both playing *h*) leads to symmetrical “injury” and the *loss* of 2 units of the “evolutionary fitness” of each player’s chosen behavior.

TABLE 1

Hawk-dove game, with payoffs expressed in units of “evolutionary fitness.” The first number in each cell of the table is the payoff for Player 1; the second, after the comma, is the payoff for Player 2.
Examples: if both players play *h*, each loses two units; if Player 1 plays *h* while Player 2 plays *d*, Player 1 gains two units, while Player 2 neither gains nor loses.

		Player 2	
		<i>h</i>	<i>d</i>
Player 1	<i>h</i>	-2,-2	2,0
	<i>D</i>	0,2	1,1

9 Game theorists were plagued by two crippling theoretical problems: 1) indeterminacy, caused by the existence of too many equilibria in their games (which made prediction impossible); and 2) the inordinate and wholly unrealistic “amount” of commonly held beliefs that their players *had* to have before *any* “solution” to the games under study could be found. EvGT cut through (2) by assuming idiotic agents who simply mimick the successful behaviors of others, and partially solved (1) by showing that some of Game Theory’s equilibria become “extinct.” For a history and non-technical introduction to EvGT see Chapter 6 of Hargreaves-Heap and Varoufakis, 2004.

By “evolutionary fitness” we mean, simply, the prospect that one’s behavior will spread in the population; the more of these units of payoffs a player amasses by adopting behavior (or strategy) i the greater the chances that his choice of i will “infect” the rest of the population.¹⁰ In contrast, mutual caution (*i.e.*, both playing d) means that each player receives 1 unit of evolutionary fitness. However, when one of the two opts for the aggressive stance, while the other behaves cautiously, the former receives 2 evolutionary fitness points, leaving the latter with neither evolutionary gains nor losses.

The importance of this game cannot be overstated. Most socio-economic interactions mix mutual benefit with a measure of conflict. For instance, when two people trade, there is often more than one price at which both will benefit. The high end of the range favors the seller, while the lower advantages the buyer. So, when they settle on a price and trade, they unlock a mutual benefit and resolve a potential conflict.

More generally, the hawk–dove game captures a broad class of social interactions in which the motivation to drive a hard bargain coexists with a fear of a breakdown of negotiations: from union–employer negotiations, to couples allocating household chores between themselves, to the endless talks at the World Trade Organization.

Let us now see how systematic discrimination is the so-called “evolutionary equilibrium” of this game. Suppose that players are drawn from a large population and are engaged in such contests repeatedly and against fresh opponents each time. EvGT focuses on the evolution of behaviors h and d and studies the way these evolve in response to their relative “success.” The evolutionary idea here is that players are, somehow, *programmed* to choose a behavior at any point in time but, and this is the rub, also that behaviors get “copied” by other agents in proportion to the payoffs they yield relative to average payoffs in the population.

Suppose that the population is purely homogeneous (*i.e.*, players are perfectly identical and thus indistinguishable from one another). Then there exists a unique evolutionary equilibrium according to

¹⁰ In these models, one’s behavior spreads within the given population either through mimicry or through having offspring who copy one’s behavior. The idea here is that an injury, following the outcome hh , reduces the chances that others will copy behavior h (since it is not successful) or that one will reproduce successfully.

which one third of players will be acting like hawks and the rest will be playing cautiously (like doves). Let's prove this simply:

Suppose that everyone were programmed to behave cautiously (play d). In each interaction each player receives payoff 1. Suppose now that, for some unspecified reason known in biology as a *mutation*, some player switches to strategy h . It might be the result of an accidental mistake, the arrival of a weird "stranger," an experiment with an alternative, off-beat, behavior, etc. Whatever the reason, the lone player selecting behavior h in a population of doves will thrive, collecting payoff 2 in each interaction (with her opponent collecting 0). If this relative "success" translates into a spreading of her mutant behavior within the population, hawkish behavior will become increasingly prevalent. Will it take over the whole population, thus turning a population of doves into a population of hawks?

No! For if all players were to behave hawkishly (play h), everyone would be receiving payoff -2 in each interaction and, therefore, some mutant dove (*i.e.*, a lone d -playing agent) would be benefiting more than the rest. Thus, a situation where all players are doves or all are hawks cannot be sustained. In the former case the proportion of doves will fall while in the latter it will rise. When will an evolutionary equilibrium be achieved? It is easy to show that the answer is: when *exactly* one-third of the population behave aggressively.¹¹

The above result rests on the assumption of zero variety in the population. No variety, no genuine evolution! This is why evolutionary theorists reject the zero variety assumption as unsustainable, both in nature and in society. Even identical twins, they say, develop some distinguishing features as time goes by. The question is: Can such arbitrary and meaningless distinctive features bring about a difference in social and economic outcomes? The answer is affirmative.

11 *Proof:* All players are identical. When a player chooses behavior h she will either receive payoff -2 (if the opponent also plays h) or payoff 2 (if the opponent plays d). Let p be the proportion of players choosing behavior h . Then, the probability that one's opponent will choose h equals p and that she will choose d equals $1 - p$. So, a player choosing h will receive payoff -2 with probability p and 2 with probability $1 - p$. On average, a hawk will collect payoff $-2p + 2(1 - p) = 2 - 4p$. Similarly, a dove will collect either payoff 0, with probability p , or payoff 1 with probability $1 - p$. On average, then, a dove collects $0p + 1(1 - p) = 1 - p$. We have assumed that hawkish behavior spreads in the population when behavior h is more successful, on average, than behavior d . Algebraically this happens when $2 - 4p > 1 - p$, or $p < 1/3$. In words, when the proportion of "hawks" is less than one-third, this proportion will rise. And when it is greater than one-third, it will fall. Ergo, it will stabilize when it is exactly equal to one-third; that is, $p = 1/3$ is the game's evolutionary equilibrium when players are wholly indistinguishable from one another. QED.

Let us suppose that, beginning with a homogeneous population, the passage of time introduces some tiny, arbitrary and meaningless differentiation that is observed by the players. For simplicity, suppose that this distinction segments the population into two types of player; *e.g.*, those with slightly longer hair than others, or birds sporting feathers of one of two colors, etc. All along, we maintain the assumption that, despite the newly observable distinctive feature, players remain identical in all other (“serious”) respects: they have exactly the same talents, prowess, resourcefulness, rationality, etc.

The original question reappears: Will this meaningless variety in features affect the social outcome? EvGT proves it will but, also, that in an antagonistic game like the hawk–dove game a systematic pattern of social discrimination will emerge and persist. The proof of this theorem is simple. Suppose there are two types of otherwise identical players: one with meaningless but observable feature “red” and the other with feature “blue.” In her interactions with opponents, a blue player can now distinguish between *two* observed frequencies: the frequency with which blue players like her behave aggressively (h) and the frequency with which red players do so. Let the two frequencies be denoted by β and ρ , respectively.

At the outset, since we assumed that we begin at a point when no distinctions are visible, the evolutionary equilibrium is one where, on average, $\beta = \rho = 1/3$; that is, on average, one-third of the blue and one third of the red subpopulations act aggressively because, say, players are indistinguishable from one another. No variety among players, no discrimination! At this point, suppose that their arbitrary color becomes noticeable. Only by a freakish accident will the observed frequencies of β and ρ be exactly equal to one-third. Just like it is improbable that a thousand throws of a fair coin will result in *exactly* 500 heads and 500 tails, equally here it is almost impossible (especially in large populations) that $\beta = \rho$. More likely, either β will exceed one-third by a tiny amount, and ρ will be less than one-third by an equally small amount, or vice versa.

Let us take, for illustration purposes, the case where $\beta > 1/3 > \rho$. As the frequency of aggressive behavior by blue players is slightly greater than one third, red players who interact with blue players will have good cause to behave cautiously; that is, choose d (see note 11, where we proved that, if a player’s opponent will choose h with a probability greater than $1/3$, she gets higher average payoffs by adopting

behavior d). By the same token, the blue players who are facing red opponents benefit from acting aggressively. But this means that because β was slightly greater than ρ at the outset, it will start rising further while ρ is falling accordingly! And vice versa.

In short, the population is, at the outset, balancing precariously on a knife's edge. An initial random difference in the behavior of reds and blues pushes the population in one of two directions: Either the reds will start acting increasingly aggressively, while the blues are drawn to a more cautious stance, or the opposite. Once the proverbial ball starts rolling, it will go all the way until it reaches one of two evolutionary equilibria: Either the reds will dominate the blues, or vice versa.

Thus, a theory of systematic discrimination within a population of virtually identical automata has come into being. Naturally, it cannot predict which group will dominate. What it does predict is that a homogeneous population has a natural tendency to subdivide (along the lines of arbitrary distinctions) and that *some* group will emerge that dominates the rest as long as the structure of the socioeconomic interaction is antagonistic (*e.g.*, of the hawk–dove variety). In summary, EvGT helps clarify how:

- (2a) extraneous characteristics can “seed” conventions that advantage one type of stimulus-response-driven individual relative to another (even if the difference across individuals is arbitrary); *and*
- (2b) the resulting conventional discrimination is stable because of the fact that behaviors that eschew the convention have a tendency to become extinct.

How significant is this for social science? Do the above results hold when the agent is a thinking, rational human being, as opposed to some pre-programmed bird or automaton?

An affirmative answer to the last question is given in Hargreaves-Heap and Varoufakis, 2002. This work reports on an experiment designed to test the above theory under controlled laboratory conditions. Experimental subjects (mostly university students) were placed in a computer-networked lab where they played hawk–dove repeatedly, with payoffs taking the form of dollars (rather than evolutionary brownie points).

To test EvGT's propositions (2a) and (2b) above in a human context, the experiment was run in two different formats: The first

was the *control treatment*, in which the games were played under conditions of complete anonymity. Subjects simply had no information whatsoever regarding their opponent/partner. Thus anonymity simulated an environment in which subjects cannot distinguish between their partners/opponents, thus rendering the population homogeneous at the level of individual perception. The second treatment tested propositions (2a) and (2b) directly, by giving players a single piece of clearly extraneous information about their opponent/partner. What was this piece of information? And how are we so sure it was extraneous?

At the beginning of each of these sessions, players picked a card at random from a pile of cards, half of which were blue and half red. Thus each player's "color" was determined. Once the sessions began, and in each round, subjects were informed of the color of their partner/opponent. Naturally, this information was as extraneous as it could have been: everyone knew that, since it was commonly known that the color assignment was random, it conveyed no significant information regarding their partner/opponent's character. The question then became: Would propositions (2a) and (2b) above be confirmed in this experimental setting?

What precisely would confirm them? They would be confirmed by the observation of significantly different degrees of aggression in cross-color meetings between the blue and the red subjects. This is precisely what was observed: in about ten rounds or so, one of the two colors had come to dominate the other. Subjects of one of the two colors evolved a tendency to act more aggressively toward subjects of the other color (than to subjects with the same color as themselves). In some sessions it was the blue players that dominated the red, while in others the reverse was observed. Moreover, subjects with the color that evolved as "inferior" developed a tendency to submit to the enhanced aggression of their differently colored opponents by adopting a far more cautious approach to them (compared to the average incidence of cautious behavior observed in experiments without any color assignments). At the end of each session, the dominant color group ended up with at least three times the money of the other.

In short, it seems that EvGT's results regarding the evolution of arbitrary discrimination among automata extend nicely to human behavior (at least under laboratory conditions). Why should the student of history care? Because, I suggest, these results confirm the

suspicion that *discrimination is the result of evolved institutions that distribute social power in ways having little or nothing to do with personal characteristics, aptitude or application*. They resonate with the progressive view within the social sciences (*e.g.*, de Tocqueville, Marx, Foucault, etc.) that the secret of systematic oppression lies more in the mind of the oppressed than in the mechanisms of oppression consciously devised by the oppressors.

The above is of great importance to those (*e.g.*, feminists, anti-racists) striving to argue that the cause of observed inequities in the distribution of social roles, wealth, income, status, etc. is to be located in the *structure of social interaction*, rather than explained with reference to differences in human capital, aptitude, application, DNA, and so forth. Of course, this does not imply that all hierarchies reflect nothing but arbitrary differences in appearance. What it *does* show is that discrimination among different strata, groups, classes, etc. can evolve in a manner that is radically uncorrelated with the personal characteristics of the individuals involved.

Indeed, if patterns of highly differential income distributions, and robust discriminatory conventions, emerge within 45 minutes in our experimental laboratory on the basis of a random color assignment), what should we expect of emotively charged bodily differences (such as different reproductive systems, skin color, etc.) in societies with thousands of years of history?

Having established that EvGT has an interesting story to tell regarding the evolution of divisions, discrimination, and inequality, I now turn to the obvious question: What is the implication of these results regarding history and the emergence of class societies?

3. An Evolutionary Account of History, Part 1: The Joint Evolution of Social Power and Ideology in Primitive Societies

The previous section can be interpreted as a sketch of EvGT's explanation of evolving behavior in populations of antagonistic agents. The analysis focused on the hawk-dove game as paradigmatic of populations that adapt their behaviors in the context of contests over the distribution of benefits (prey, shelter, partners, etc). Before the analysis can be applied to even primitive human societies (*e.g.*, hunter-gatherer communities), it needs to be elevated to a significantly higher level of analytical complexity. This I attempt below.

Recall the reported experiment of the previous section. It suggests that EvGT's results apply equally to humans and animals: discriminatory practices evolve which achieve the minimization of conflict¹² through i) the arbitrary division of agents between groups; and ii) a pattern of dominance by one group over others. Yet nothing in the analysis itself is, so far, specific to human societies. Social regulation *requires a substantive theory of the ideology and institutions of social stratification.*

To begin discussing the social institutions that “make a society,” it is important to define the term. This paper defines a *primitive institution* as *any mapping from individual motivation to social outcomes that cannot be reduced to data on private motives and constraints.* In this sense, the socioeconomic equilibria of the previous section appear as primitive institutions, but only to the extent that the practice of, say, the reds conceding in hawk–dove, while the blues take the spoils, is sustained by the tendency of “deviant” behaviors (behaviors that deviate from the evolved behavioral pattern) to fade away, victims of evolutionary pressure.

To move beyond primitive institutions, and toward fully social ones, the theory must be able to encompass the idea of *convention*, in the sense of Lewis (1969). Central to this idea is the concept of *self-fulfilling forecasts*. In our laboratory, for instance, the dominance of one group over the other was sustained by the participants' *forecast* that this is what would happen. Such predictions became self-fulfilling because, once they were shared, no individual could benefit by acting in a manner that contradicted them.

This is the first analytical step in the transition from primitive to substantive social institutions: for the theory to throw a bridge from the animal republic over to the human world it needs to introduce something specific to human motivation: a *system of beliefs*! In the case of ants and bees, adaptive behavior is all evolutionary biologists require to explain the evolutionary dynamics; any talk of institutions or convention is surplus to requirements. However, when the players

¹² Notice that the discriminatory conventions in hawk–dove eliminate (in meetings between members of different groups) the possibility of conflict; that is, of outcome *hh*. The reason, of course, is that when one group dominates the other, and two agents meet belonging to different groups, one of them will submit to the other (choosing *d*) while the other will act aggressively (*h*). The result is a “clean” outcome which involves no chance of losses for either agent. Absolute discrimination and ruthless social dominance, however vile, at least regulate behavior in such a way as to eliminate conflict!

are human, the evolution of behavior is underpinned by (and gives rise to) an evolving belief system. The latter is, in turn, equivalent to a fully specified *social institution*, consisting of *mutually reinforcing practices, predictions and normative beliefs*. In short, an *ideology*!

Returning to our experiments as a source of examples, consider what happened after the dominance of one group (say, the blues) over the other (say, the reds) was established: in cross-color meetings the blues *expected* the reds to concede (choose *d*) while they acted aggressively (behavior *h*); an expectation that was confirmed by actual behavior. On one occasion, however, I mischievously programmed the computer network to lie to the blues, informing them (falsely) that their red opponent had "rebelled" (by choosing aggressive behavior *h*). I was simply curious to see if *anything* would happen.

It did! The blues reacted with a sharp, emotional response; they shouted (in an otherwise tranquil computer laboratory) abuse at unseen deviant red players, displaying a fascinating form of *moral outrage* at the "subservient" opponent's audacity! A remarkable transformation had, thus, occurred. Only an hour or so before, these players had entered the laboratory as complete strangers. They were given a color code at random (blue or red) and were seated in cubicles that prevented them from even seeing each other. Within the space of half an hour or so, an asymmetrical institution had evolved, sustained solely by the *power of their prophesy*.

The evolved institution regulated behavior, minimizing conflict and, effectively, granting "property rights" to the players of one color (who could increasingly bank on getting \$2 every time they met a player with the opposite color). Before too long, however, the prophesy-reliant institution acquired an ideological, perhaps a moral, dimension: *the dominant players had begun to believe not only that they would get the \$2 but, amazingly, that they also deserved this outcome!*

This type of transformation is only possible in human societies: behavioral patterns beget predictions which, in turn, beget normative beliefs that reinforce (and sometimes undermine) the established conventions. Why does this occur? An interesting answer comes from David Hume's explanation of how mere conventions *annex virtue to themselves and thus become social norms, or norms of "justice"* (Hume, 1988). Cast in modern terms, the idea is that a community's institutions become more resistant to deviant behavior ("mutations," in the parlance of evolutionary theory) when people not only expect others to

behave in accordance with the established conventions but, also, feel that deviating from them is somewhat — *wrong*; even *morally defective*.¹³

Sugden (1986; 1989) expands on this theme with the point that, as conventions begin to impart “moral” beliefs, they gather additional resistance to behavioral “mutations.” Put simply when, in addition to making us *predict* that we *shall* all adopt some behavior, a convention also convinces us that we *ought* to adopt it, it proves far less susceptible to deviant behavior. And since robust conventions minimize conflict and enhance benefits *on average*, morality is an illusion *functional* to the *average* individual’s petty interests.

Moreover, when a ruthlessly discriminating convention emerges (as in our experiment), people find it hard to accept that the convention is in some sense arbitrary while also being so discriminatory. So people remove the resulting *cognitive dissonance* by finding, or inventing, additional principles that will justify the actual convention because it is “just,” “fair,” etc. When they succeed in this, the convention becomes more entrenched as both its beneficiaries *and those it discriminates against* are less likely to contravene it.

In summary, an interesting theory of the primitive formation of ideology obtains from this convergence of EvGT and neo-Humean thought (see also Binmore, 1998). Morality is depicted as the reification of conventions whose *raison d’être* is to coordinate behaviors to some equilibrium devoid of waste and conflict. In this context, norms of justice are cast in the same light: as conventions that imbue people with expectations of what is right, just, or wrong.

At the political level, this conversion of predictions to ethical beliefs gives rise to the ideological notion of the “common good,” which is, in this account, another illusion underpinned by the fact that convention-following brings greater *average* benefits (unequally, of course), when compared with what would come to pass in the absence of *some* convention. Meanwhile, *at the level of the individual*, as we saw above, the dominant ideology infects the thoughts not only

13 Hume insisted that we learn not only to *predict* that others will follow the established convention but, additionally, that we *expect* them to do so. Indeed, when they fail to do so, many of us are often filled with moral indignation at behavior “prejudicial to human society.” At that juncture, our predictions *vis-à-vis* others’ behavior have become *normative, or moral, expectations*. In Hume’s own words, at some point of the evolutionary path, the “is” and the “will” become a “must” or an “ought”: “. . . when of a sudden I am surprised to find, that instead of the usual copulations of propositions, *is* and *is not*, I meet with no proposition that is not connected with an *ought* or an *ought not*.”

of the agents who benefit the most but also of those who benefit the least. A simple form of what Marx referred to as *false consciousness* is, therefore, given an evolutionary explanation consistent with the observation that, after centuries of discrimination, many women feel that men *deserve* the leading social roles, or that most men in Papua New Guinea accept the *moral superiority* of white, male, American Protestant preachers.

Let us take stock. The preceding evolutionary theory of social institutions offers five insights:

- (3a) When the interaction distributing benefits in a population is asymmetrical, institutions evolve that divide the population along rigid lines of stratification with some groups profiting at the expense of others.
- (3b) The lines of social division, and the composition of the dominant and subservient groups, are drawn arbitrarily and independently of the motives, talents, powers or aptitude of their members.
- (3c) As the iniquitous primitive institutions evolve, the resulting divisions have a tendency to subdivide and multiply further, thus creating institutional discrimination *within* the major social strata they have generated at an earlier stage of the evolutionary process.¹⁴
- (3d) The divisions give rise to, and are reinforced by, suitable normative beliefs forming an ideology that reflects the underlying discriminatory practices.
- (3e) The evolved conventions of distributing assets and roles asymmetrically spread from one realm (or interaction) to another *by analogy* (Sugden, 1986; 1989).

This last point deserves some elucidation. The previous section focused on one simple interaction, the hawk-dove game. Societies are, of course, founded on many different types of intercourse. Some are antagonistic, like hawk-dove; others take the form of *prisoner's dilemmas*¹⁵; some pit an urge to coordinate one's actions with those

14 The point here is that just as a homogeneous population tends to subdivide (see previous section), so the resulting subdivisions have a tendency to subdivide further. This "deepening" of divisions renders the overall population's behavioral equilibrium more stable. For an analysis, see Hargreaves-Heap and Varoufakis, 2004, Section 6.3.3 of Chapter 6.

15 Agents may be caught in a prisoner's dilemma when they have a choice between cooperating with each other and acting selfishly. Suppose that mutual selfish behavior makes all agents worse off than mutual cooperation. Suppose, however, that the highest private benefits are to be had when one acts selfishly against cooperative opponents (and the lowest benefits accrue to a cooperator who is betrayed by the selfish behavior of others). In this situation (assuming it is to be "played" out only once by agents who care only about their

of others against an opposite urge to go it alone. Point (3e) refers to the evolutionary argument that conventions, once established in the context of one type of interaction, “migrate” to other interactions.

Take, for example, hunter-gatherers operating cooperatively in order to catch large prey (*e.g.*, stags). Success depends, as initially pointed out by Rousseau (1762), on their capacity to develop resistance to the centrifugal forces of prisoner’s dilemma-like urges that are best kept in check by conventions for dividing the spoils around the campfire.¹⁶ In other settings (*e.g.*, areas where the prey migrates or the weather conditions change rapidly from one season to the next), hunter-gatherers must nurture nomadic conventions for both hunting and distributive purposes. In large areas with scarce, small prey, hunter-gatherers are more likely to work alone. However, they are still likely to come up against one another and compete over the same prey or resource, not unlike in the hawk-dove interaction which we looked at in detail in the previous section.

The socioeconomic context of hunter-gathering just described involves a melange of different interactions, only one of which is captured reasonably well by hawk-dove. In more complex settings, in which cooperative and non-cooperative interactions alternate or coexist, point (3e) corresponds to Sugden’s hypothesis that institutions spread across the different interactions that constitute a community’s social life (1986; 1989). Recent experimental evidence seems to confirm that hypothesis.¹⁷

own net private gains), each agent chooses the selfish act (since it is best to act selfishly *independently of what others do*) and, as a result, they all suffer. This game is particularly useful for capturing the difficulty of organizing *collective action*. For though the latter may be beneficial to all, each may have a good cause to free-ride on the proverbial back of others.

16 J.-J. Rousseau (1762) was concerned with the problem of collective production where each member of a community of producers must choose between different degrees of commitment to a common good whose eventual value will be proportional to the effort of the *least committed team members* (who can, thus, “let the side down”). Rousseau’s team of hunters could either join forces in order to catch a stag, so that all can eat well (a feat depending on the commitment to the task of each and every member; as opposed to the average commitment), or abscond and hunt separately for smaller prey (*e.g.*, rabbits) to be eaten individually. His conclusion: it takes social optimism and public spiritedness to bring about the socially optimal outcome (the stag’s capture).

17 In Hargreaves-Heap and Varoufakis (2002), for example, when a third co-operative choice was introduced half-way into the experimental session, the discriminatory convention that had previously evolved in the context of hawk-dove metamorphosed into a fascinating new practice: members of the subservient group started cooperating with one another at a rate of 90% whereas members of the dominant group never cooperated. Moreover,

Returning to the attempt to weave a history of human society based on EvGT, a good starting point is the presumption that social institutions, and hierarchies, had already evolved even before we were “fully” human. As geographical and climatological conditions necessitated more cooperative patterns of primitive production (*e.g.*, nomadic or collective hunting), these hierarchical conventions spread by analogy from the realm of hawk-dove-like contests to the ways and means by which collective produce was privately appropriated. To the extent that the community’s evolutionary fitness was intimately linked to the solidity of those conventions, developments that weakened any tendencies to “disobey” the established conventions were reinforced. Of all such developments, the rise of the ideology of stratified societies (including concomitant religious beliefs) was the most effective.

The evolutionary fitness of the institutions of power and belief was improved further by two separate developments: First, the subdivision of populations into sub-strata entrenched the conventions of discrimination, by ensuring that a significant minority of those belonging to a main subservient group enjoyed certain privileges relatively to specific sub-groups (see point 3c above). Second, the evolution of human language, around 100,000 years ago, which facilitated, through the invention of moral signifiers, the emergence of concomitant ethical beliefs that “enabled” people to feel not only that the violation of given conventions is dangerous but that, more poignantly, it is also *morally problematic* (recall Hume’s ironic point, in note 13, of the ease with which our language slips from “is” to “ought” statements.)

That EvGT can provide an interesting narrative on the emergence of social power and ideology in hunter-gatherer societies is, thus, not in dispute. The question is whether it has a sophisticated story to tell about societies with collective production, ranging from the agrarian to the capitalist.

4. An Evolutionary Account of History, Part 2: The Path to Capitalism

Humanity’s Great Leap Forward came with the development of farming, which put us on the path of socialized production (a pre-

recent unpublished experiments (conducted by this author) show that once a pattern of dominance is established in simple accumulative contests of the hawk-dove variety, it colonizes the ensuing, more complex bargaining contexts.

requisite for sustainable farming practices), organized armies (for the protection and/or appropriation of stockpiled food), bureaucracies (for the organization of collective effort and the distribution of the resulting surplus), writing (for purposes of bookkeeping), the evolution of differential resistance to new diseases (leading to the genocide of those without it by those with it; *e.g.*, Native Americans and Aboriginal Australians), the technological progress that led to greater capacities to create (*e.g.*, metal technology for the manufacture of ploughs) as well as to destroy (technological advances in the development of weaponry), etc. However, even before we embarked collectively down that path, we came to it fully equipped with institutions founded upon the discriminating conventions developed at the earlier, hunting-gathering stage of socioeconomic development. Let us see what the EvGT-based narrative of the previous sections can contribute in this regard.

Anthropological fieldwork tells us that the hierarchical norms of dividing contemporary goods and chores did not begin with socialized food production. As the latter did not replace hunting-gathering abruptly, but coexisted with it for centuries (see Diamond, 1996), underneath the surface of the norms of surplus distribution there lay hidden many layers of prior discriminatory conventions stemming from an earlier hunting-gathering era. By simple deduction, the norms that determined who controlled the land must have fed into new, analogous norms regarding control of the surplus.

Crucially, the moment food production comes into the picture, the epicenter of social power shifts from appropriation-*cum*-consumption to control over the production process. The simpler institutions of primitive appropriation can hardly carry the burden of this major socioeconomic transformation. Rituals and norms for dividing spoils and determining hierarchies around the campfire are one thing; rules governing access to land, the division of labor between farmhands, smiths, priests and soldiers, etc. are quite another. Thus, the nature of conventions "mutated" a second time: After having acquired a moral dimension (and transformed into an ideology), the social conventions of distribution crystallized into a new form of written laws, complete with the state authority to enforce them. Once agricultural production appeared on the scene, "legitimized" highly divisive institutional power became the natural attractor of the socioeconomic process.

While state coercion played a significant role in the equilibration of the evolving social economy, it was scarcely enough. Brute force could not, by itself, regulate the increasingly complex relations of production and distribution. The evolved social hierarchies increasingly relied on the exercise of a subtle, subterranean power: the power to extract surplus from others without constant resort to naked violence. I shall refer to this subtle power to control surplus production as *extractive power* (see C. B. McPherson, 1973, for the original articulation of this notion).

Extractive power is erected on a solid foundation of conventions, norms and ideology. Below I offer a re-worked definition of it in accordance with the preceding analysis.

Generally speaking, person *i* exercises *extractive power* over *j* if:

- (4a) *i* and *j* are virtually identical except that *i* sports extraneous feature *F* which places her in the advantaged social group *A*, leaving *j* in disadvantaged group *D*.
- (4b) *i* can "persuade" *j* to perform task *T* which results in surplus *S*.¹⁸
- (4c) *i* can, courtesy of her *A*-group membership, enforce property rights over *x*% of *S*.
- (4d) *j* would not have performed task *T* for $(1 - x)\%$ of *S* had the distinction between group *D* and *A* not evolved previously.
- (4e) Social norms prevail upon *i* and *j* to think of the $[x, (1 - x)]\%$ distribution as "fair."

Extractive power, as defined above, is thus a straightforward extension of asymmetric conventions for distribution of non-produced goods to a community which produces assets in the context of collective manufacture. In principle, it can emerge in hunter-gatherer communities too; in the sense that some group may develop, theoretically, a capacity to compel others to hunt/gather on their behalf. However, such conventions are less likely to take hold and command a significant proportion of work effort when individuals have the opportunity

¹⁸ Tony Aspromourgos, in a private communication, argued, quite rightly, that in most systems of socialized production the individual worker's contribution to the surplus, *S*, is not only unobservable but, in a sense, ill-defined. In fact his argument adds to my point in the next section, where I argue that capitalism (due to its much greater reliance on collective or social production) both a) enhanced extractive power (by making individually produced surplus invisible and ill-defined) and b) depended crucially on wage labor, since there are natural limits to the extent that laborers can become independent contractors.

to abscond and fend for themselves. The more restrictive the access to productive resources (*e.g.*, the more fences there are around fertile land) the greater the preponderance of extractive power.

In this account, social strata that gained conventional control over scarce land acquired conventional control also over others' productive efforts. Extractive power became, in this manner, inextricably linked to the technology of production and the *outside options* of individuals belonging to groups devoid of extractive power. Thus, *social classes* developed out of strata. The power to compel, under the definition above, is not the form of power associated with brute force but the subtler type of power that relies on making offers that the "other" cannot refuse, due to a paucity of viable outside options. When this differential extractive power is founded on group membership, which in turn revolves around control of productive means, a class society becomes fully operational.

However, the divisions spread well beyond and across the class divide. As argued in the previous section [recall point (3c)], the forces forging evolutionary stability within divided societies give rise to a rich pattern of discrimination that cuts through and across the line separating dominant and subservient strata and classes. Divisions, in short, *subdivide and multiply* in response to evolutionary social pressures. This is a crucial point for progressive social theory, which has often slid toward pointless discussions on whether class is more important than gender, gender than race or race than class. EvGT helps us dismiss this debate altogether by demonstrating that sustainable discrimination cannot be built on binary distinctions.

Of course, the sustainability of social divisions is achieved by a combination of the subtler type of power defined above and brute force (often taking the form of state power). The point of the present evolutionary account is twofold: First, that societies relying more on multidimensional extractive power and less on violence for the regulation of production must have displayed greater evolutionary fitness than the rest. Second, that violence (privately or communally dispensed) probably worked well only intermittently, whenever the normative beliefs and ideology behind extractive power wavered.

Returning to the centrality of *variety* in any evolutionary theory, the possibility of coexisting and inter-weaving patterns of extractive power allow for the possibility of older and newer conventions to operate side by side within the same institutional framework; at least

for a while, blurring further the distinction between the dominant and subservient groups. The group privileged by history in the land-distribution game (the landed aristocracy) became a *social class-for-itself* once: i) its privileges became hereditary (and the group could reproduce itself as a group); and ii) it embellished its extractive power over the rest with moral meaning (a dominant ideology).

The dominant ideology was subsequently reinforced by the complexity of the conventions according to which control of the land and its output was dispensed, and by the normative beliefs in which they were disguised. History tells us that, when the exercise of extractive power became too obvious, revolt beckoned and, quite often, the heads of the dominant group's members rolled. In short, the greatest defense of the conventions of the agrarian societies was the capacity of norms founded on extractive power over the produced surplus to become invisible.

So, it took at least two great transformations to solidify, stabilize, and massively boost extractive power: *feudalism* and, then, the advent of *capitalism*. While a theory of transition to capitalism is outside this paper's scope, it is possible to offer some thoughts, consistent with the preceding analysis, on the fundamental changes that capitalism brought to the exercise of social power, and to the nature of extractive power in particular. Capitalism took this capacity to hide extractive power to dizzying heights. Its chief achievement was to combine unprecedented inequality with formal liberty. This "paradox" sits well with the idea above that because the conventions underpinning capitalist relations of production were considerably more oblique, the normative beliefs accompanying them proved more stable.¹⁹

The question of course is: How did capitalist conventions become less visible than those they replaced? Farming, I argued above, introduced extractive power by reversing the timing of extraction.²⁰ Rather

19 A choice between hunger and selling one's labor to the highest bidder is not an easy one. But it is still a choice, compared to the "choice" between being stabbed and handing over a part of one's output to the Lord's Sheriff.

20 A reader of an earlier version of this paper protested that workers, in reality, are paid after they work. This is not the point. The issue here is that capitalists are contracted to pay wages in the short term (*e.g.*, at the end of the working week) *independently of the timing of the output's market sale*. The same reader protested that feudal lords can also be thought of as paying wages in advance, in the form of goods (such as breakfast) that they made available to peasants before the work was done and the harvest was in. Be that as it may, just as it is always true that precapitalist societies featured markets, without being market societies, equally, some portion of the feudal surplus was advanced to peasants without, however, altering the fact that the bulk of the surplus was distributed *ex post*.

than receiving a proportion of the produced goods determined after the harvest came in, the members of the socially dominant group would advance a fixed amount to the workers and claim the residual. This reversal turned the economically powerful into *residual claimants* and was intimately linked with the *commodification* of domestic labor and land which was, in turn, combined with the *forced expropriation* of people overseas (*i.e.*, slavery). These three related developments resulted in the industrial revolution, its wonders and catastrophies. Why is the timing reversal, and the conversion of landlords (and later capitalists) into residual claimants, evolutionarily significant? What does it matter who gets their share of the surplus first and who retains the residual?

Paul Samuelson, the Nobel Prize-winning economist who wrote perhaps the most influential postwar economics textbook, once famously claimed that it matters not at all; that who pays whom in the production process (the capitalists paying the workers or vice versa) is irrelevant. The reason he is wrong is twofold: The best rehearsed explanation is that, having laid out a fixed amount to the workers at the outset, capitalists acquire an incentive to squeeze as much produce out of them in the ensuing production process. A second explanation, which receives little attention, concerns the pivotal role this reversal played in disguising the social conventions at work.

Under precapitalist social relations of production, control over production largely remained in the hands of the producers. It was only after the crop came in that the distributional conventions would kick in; a fact that made *obvious* the evolved and utterly arbitrary extractive power that the owners of land had over the non-owners. But under capitalism, the temporal reversal of residual claims meant that workers lost control over the production process. For the first time in human history the residual claimants paid in advance for the privilege of exercising their extractive power. Given the inherent risks of paying for something in advance, the task of removing the cognitive dissonance resulting from the preposterous social asymmetries that capitalism brought to the fore was eased substantially.

Those privileged by the new capitalist conventions could legitimize their gains based on the mythical notion of profit as a just reward for risk-taking. More importantly, those disadvantaged by the same conventions could live with their situation more easily, given a combination of normative beliefs shaped by: i) the seemingly symmetrical

position of capital and labor ("we receive profit in return for laying out in advance our capital, and you receive this capital in advance in return for your labor"); and ii) the soothing impact of formal liberty for all.

In conclusion, EvGT's single most important insight into the nature of capitalism concerns the mechanism by which the latter consigns the very extractive processes that it relies upon to near-perfect invisibility. This, in turn, throws useful light on:

- (4α) the deepening antagonistic character of the games we play as technology makes it easier for telephonists in India to take emergency calls from Colorado and production to be shifted at a moment's notice in search of the lowest wage rate.
- (4β) the increasing fragmentation of the dominant ideology into post-modern, localized ideologies that lack some "common currency" (*e.g.*, the demise of the Enlightenment ideals of liberal society) and, more importantly, make it hard for us to distinguish the overarching socio-economic system's structure.
- (4γ) the increasing tendency of capitalism to obfuscate the essence of distribution by altering the timing of payments and delivery of goods (*i.e.*, the creation of futures markets that require a great deal of technical expertise to disentangle).
- (4δ) the ensuing crisis of the state, whose authority is undermined both by (4α) and (4β) above.

5. *History Versus Evolution, Part 1: The Impasse of Evolutionary Approaches*

This section argues that a) the evolutionary approach to human history is, *by its very nature*, insufficiently evolutionary, and that b) it will always fail, *by its own criteria*, to offer a substantive theory of capitalism.

The reason for this bleak prognosis is that evolutionary theory, if it is to offer determinate models, can only portray the generation of variety in a manner that *must* be too restrictive to allow for genuine historical change. The key issue here is the way evolutionism is constrained to model "subversive" behavior as apolitical, random "mutations."

Recall EvGT's account of discriminatory institutions which correspond to evolved asymmetrical social conventions. Their stability (and thus theoretical power) relied on the observation that they could

not be undermined through individual action ("mutations"). However, collective (that is, political) action could readily subvert them, as it has been doing throughout history (since, at least, the days of Spartacus). Yet, evolutionary theory can have nothing to say on this.

This is not due to neglect but, rather, to theoretical prudence: for if we allow individual mutations to be cointegrated with past behavior, as well as with the behavior of others, we shall end up with an infinity of predictions. Radical indeterminacy may sound like a celebration of politics as the practice of shaping a society's mutation probabilities and, eventually, of the "game." But, since a theory that explains *all* possible histories as consistent with the evolutionary approach is a theory with very little explanatory power, it also ends all hope that evolutionary theory holds the key to understanding human history (confining it instead to a study of the properties of the various, potential "equilibria"; see Mailath, 1998).

To their credit, a number of evolutionary game theorists have understood this well. Foster and Young (1990), for instance, acknowledge that politics is what happens when mutations are coordinated into aggregate shocks that test the established conventions. Kandori, Mailath and Rob (1993) examine the impact of rational experimentation in finite and discrete populations. Bergin and Lipman (1996) demonstrate that allowing the mutation probabilities to depend on current behavioral codes (as opposed to being random and uncorrelated with present conventions), yields a devastating theorem (known in the trade as a *Folk Theorem*): almost *any* conventional behavior can become disestablished, and any alternative may take its place, if "mutants" coordinate their mutation probabilities appropriately and in response to the current behavioral conventions.

Summarizing this section's first criticism, the evolutionary approach loses all analytical power the moment it allows humans to do what they have been doing throughout history: to coalesce, subversively, around common goals and influence one another's deviant acts, often through dialog, so as to disestablish discriminatory conventions. Evolutionary theory's inescapable need to assume that all social "deviance" is, effectively, "apolitical" motivates this section's title. Modeling mutations as random events may be admissible in biological theories of, *e.g.*, the evolution of genes. However, when it comes to human societies, the mechanism generating social and political anti-establishment behavior cannot be treated as *statistically*

independent of the mechanism that alters our character, capacities and social relations.

Turning to this section's second claim, viz. evolutionism's special difficulties in grappling with capitalism's essence, my argument is that it is not equipped to grasp: i) the particular reliance of capitalism on manufactured belief, and ii) the process by which belief is manufactured under capitalism, as well as the manner in which it is intertwined with the technology of production responsible for both capitalism's material advances and economic crises.

Belief, Section 4 argued, has a special role under capitalism. The reversal of the timing of surplus extraction, in conjunction with the laborers' formal liberty, made capital accumulation possible. It enabled greater investment in privately controlled productive means through the mobilization of present and current surpluses across activities and sectors. However, this process required trust and, thus, it put *belief* at center stage both at the micro and the macro level.

At one level, extractive power was enhanced by the combination of technical change (made possible by futures markets for capital and commodities), the social norms of property rights and of the related division of labor under conditions of mass production, as well as the belief system that sprang from these and rendered extractive power invisible (thus, enhancing its sway and scope). At another level, the newfangled system's dynamism hinged on the maintenance of optimism among the capitalists viz. future demand for their commodities.

The first level was one that Marx explored meticulously, while our understanding of the second owes most to Keynes. From our current perspective the question is: Did a fundamental shift occur with the transition to capitalism that evolutionary theory can trace and elucidate? Was the Great Transformation, to borrow Karl Polanyi's phrase (Polanyi, 1945), a series of mere adaptations of pre-existing norms of production and distribution?

The answer depends on the theory's capacity analytically to penetrate the major forms that are central to capitalism; *e.g.*, the idea that everything is (or ought to be, at the "right" price) a commodity for sale, of formal liberty for all, of profit and wages as potentially "fair" payments for "productive" contribution, etc. Marx spent much ink on the evolution of the commodity and of capital as analytical categories that are anything but mere evolutionary adaptations of pre-capitalist entities. As commodity exchange became the exclusive

means of survival, the commodity-relation replaced human relations. Capital, *i.e.*, the manufactured means of production, “was not a thing, but a social relation between persons. . . . Property in money, means of subsistence, machinery, and the other means of production, do not yet stamp a man as a capitalist if there be wanting the correlative the wage worker” (*Capital I*, in Marx and Engels, 1979).

But if capital is a contested relation of production (as opposed to some physical “thing”), then its value is a matter determined not only by the network of conventions ruling over this relation *but also by the systemic (as opposed to random) attempts to subvert them*. While evolutionary theory has an interesting story to tell about the former it *must* (as argued above) remain mute on the latter, if it is to keep at bay the terrible specter of indeterminacy. In this context, although the evolutionary approach may have on offer various intriguing narratives on how social conventions reflect the *jointly* evolving technologies and relations of production, all such narratives are condemned to be innocent of a critical aspect of capitalist social dynamics: the inevitable, collective, political resistance of those who are, systematically, at the sharp end of extractive power.

In conclusion, steam engines, mechanical looms, and computerized robots are, at once, the secret force behind splendid productive capacity *and* the midwives of our ideology. As technology progresses, it causes ruptures in the established behavioral conventions and the associated institutions by means that evolutionary theory can only understand if it surrenders fully to theoretical impasse (that is, indeterminacy due to an infinity of equilibria). Evolutionary theory, consequently, faces a dilemma between: a) to abandon the cause of elucidating history in general and capitalism in particular, sticking instead to determinate models where variety is generated by random, apolitical, mutations; or b) to espouse history, by ditching its penchant for modeling exercises.

Interestingly, astute practitioners of social evolutionism are increasingly drawn to similar conclusions. Sugden (2001) argues that EvGT failed because, unlike biologists, its practitioners have no interest in historical data (the equivalent of fieldwork with fossils). Hodgson (2006), though an adherent of Darwin’s method, adds that an evolutionary approach to social institutions must be primarily historical, if it is to bear fruit. On a positive note, while the historical is irreducible to evolutionary *mechanics*, a thorough grasp of why that is so is an excellent start for any historical inquiry.

6. *History Versus Evolution, Part 2: Values and Liberty*

This section outlines the third, and final, major criticism of the evolutionary approach: evolutionism cannot furnish a *critique* of evolved institutions useful either to the historian or the active citizen. To illustrate, consider the question: What is *really* wrong with a world of contented slaves or, less emotively, one in which the dominant ideology has made most people accept (and even like) the institutions and norms of the prevailing social mode (feudalism, slavery, patriarchy, contemporary capitalism, etc.)?

An answer along the lines of a moral judgment invoking the notions of injustice or illiberty, and based on empirical observation of socioeconomic data, is *not* open to those who adopt an exclusively evolutionary narrative. For the latter dismisses moral judgments as quasi-illusions functional to the current conventions. The only route available to the critic (who has adopted the analysis so far) is to ground her criticism on something *outside* the evolved belief system.

Marx, for one, focused his indignation on the *inefficiency* of capitalist *social relations* (e.g., the arrangement according to which the sets of workers and of owners are, mostly, mutually exclusive) which, in his eyes, have not evolved sufficiently to take full advantage of the available technology. His critique of capitalism turns on the argument that it represents a transitory phase of human history. The notion that capitalism is efficient but unfair is dismissed angrily, replaced by the portrait of a social system which is one evolutionary stage behind the productive capacity of the machinery that it, itself, brought into being.

Due to this "evolutionary backwardness," according to Marx, capitalism wastes human resources (in the form of chronic and fluctuating unemployment), devalues humanity (by reducing our relations to commodity fetishism), restricts real liberty for most, and requires human sacrifices upon its altar (war) in order to maintain some degree of compatibility between i) what the economy can produce; and ii) what consumers have the purchasing power to absorb.

In true evolutionary spirit, Marx rejects both bourgeois *and* proletarian moralities: for they constitute the different sides of the same overarching ideology which prevents humanity from achieving its potential (see Wood, 1991). However, while espousing an evolutionary account of how we came to be enmeshed in our current social

relations and ideology, Marx maintains hope of escaping, through the power of human reason and political action, the moral relativism that comes with vulgar naturalism.

In this sense, values matter to humans because they reflect this capacity i) to cast a critical, fully rational, gaze on what we do; and ii) to subvert the rules that "ought" to govern our behavior, not merely by means of mutation-like random experiments with alternative "moralities," or codes of conduct, but also by means of *critical reflection*, *rational dialog* and the *collective acts* that follow from these. The point of the rejection of all moralisms is that they circumscribe our capacity to understand the world and, thus, to improve on it. Only, such improvement is made impossible if, along with the moralistic bathwater, we throw away our capacity for critical reflection, as evolutionism covertly invites us to do.

There is, of course, an unanswered question: How can we *criticize* our social order (slavery, feudalism, capitalism, etc.) without resorting to the normative views that have been foisted upon us by that very social order? Although the answer is not clear, the very *possibility* of an answer is that which distinguishes a good historical narrative from a competent evolutionary model; and a progressive optimist from a conservative pessimist. Evolutionary models explain critical reasoning, moral judgments, and normative beliefs (just like genotypes and phenotypes) in *functional* terms, viz. their role in facilitating our *given* interests, which we pursue within *given* interactions, and under *given* rules. Although aspects of social life are explainable in this manner, history is intolerant of so many givens. By moving beyond them, historical approaches inspire hope of liberation from our illusions without, however, pushing us into the sinister embrace of moral relativism. The Study of History, just like the immersion into Art and Music, delivers us from artifacts of our own creation (like EvGT) which, once milked for all they are worth, we must transcend.

7. Conclusion

Capitalism is an evolving system whose evolutionary fitness is enhanced substantially by the fact that its subjects can no longer see it as such. Convinced that we live, instead, in a *permanent market-place*, in which only firms, technologies and products evolve, our vision of capitalism as an evolving entity is severely circumscribed. Like the

dwellers in Plato's cave, we only see the shadows of the invisible white-hot forces forging the shifting constraints under which we live and think.

Evolutionary theory, after its recent alliance with game theory, demonstrates brilliantly how discriminatory social power springs out of the social conventions ruling over our practices and our beliefs, and is solidified best *when it remains as opaque as it is all encompassing*. Millennia ago our beliefs and practices acquired a religious-cum-ethical complexion and, later, following the advent of food production, conventions were crystallized in written legislation. Thus they became the cement that solidified the evolved social classes and hierarchies. With capitalism, the dominant class became the residual claimant over the produced surplus. Its extractive power was then concealed almost fully by discriminatory institutions which, because of their unprecedented variety and diversity, engendered a uniform totalizing ideology regarding the illusory causality between privilege and "worth" or "virtue."

However, history, this paper has argued, is irreducible to evolution. Evolutionary models are condemned either to be oversimplified, determinate abstractions (like those of EvGT) or to wallow in indeterminacy, producing an infinity of predictions. Transformative, as opposed to variational, change cannot be charted by models that understand it solely in terms of residual random variability and some filtering process that smooths the rough edges adaptively. Even in nature, evolution, infrequently but surely, *jumps* from one state to another, either exogenously (*e.g.*, due to some meteor crushing on earth) or endogenously (*e.g.*, ecological degradation). These leaps rule out intermediate steps *via* the operation of hidden constraints which only fieldwork, as opposed to theoretical modeling, can reveal. Similarly in society, "events" like the invention of food production or the transition to capitalism cannot be charted via modeling exercises. As they are not the result of some gradual hill-climbing exercise but, rather, the outcome of what happens when evolution comes up against hidden constraints, proper historical work is indispensable.

In summary, historical developments of great note (*e.g.*, the institutions of food production and capitalist relations of production) boosted handsomely the degree of extractive power of elite over non-elite groups by shrouding their social power in a veil of obfuscation. The continuum of technological and ecological change may

be history's steam engine, but what determines its path is the tendency of humans to reflect critically upon our actions and to subvert collectively the norms that, supposedly, "ought" to be ruling our behavior. The penchant for deviance is, arguably, at least as natural a product of our evolution (even if less frequent) as the tendency to adapt. Its effect is to keep conventions of social power constantly on their toes, ready to destabilize them the moment some technological or other development has upset their evolutionary fitness. Thus history moves on in leaps and bounds with various evolutionary processes hot on its trails, struggling to catch up. No genuinely historical approach can afford to leave out of its ambit a model of humans as creative agents capable of both individual contemplation and collective subversion.

Department of Economics
University of Athens
8, Pasmazoglou Street
Athens 10559
Greece
yanisv@econ.uoa.gr

REFERENCES

- Bergin, James, and Barton L. Lipman. 1996. "Evolution with State-Dependent Mutations." *Econometrica*, 64, 943–56.
- Binmore, Kenneth. 1998. *Just Playing*. Cambridge, Massachusetts: MIT Press.
- Darwin, Charles. 1859. *On the Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.
- Dawkins, Richard. 1976. *The Selfish Gene*. Oxford, England: Oxford University Press.
- Diamond, Jared. 1996. *Guns, Germs and Steel: The Fates of Human Societies*. New York: Norton.
- Foster, Dean, and H. Peyton Young. 1990. "Stochastic Evolutionary Game Dynamics." *Theoretical Population Biology*, 38, 219–32.
- Hargreaves-Heap, Shaun, and Yanis Varoufakis. 2002. "Some Experimental Evidence on the Evolution of Discrimination, Co-operation and Perceptions of Fairness." *The Economic Journal*, 112, 678–702.
- . 2004. *Game Theory: A Critical Text*. London/New York: Routledge.
- Hodgson, Geoff M. 1993. *Economics and Evolution: Bringing Life Back into Economics*. Oxford, England: Polity.
- . 2006. *Economics in the Shadows of Darwin and Marx: Essays on Institutional and Evolutionary Themes*. Aldershot, England: Edward Elgar.
- Hume, David. 1888 (1740). *Treatise of Human Nature*. Ed. L. A. Selby-Bigge. Oxford, England: Oxford University Press.

- Kandori, Michihiro, George J. Mailath, and Rafael Rob. 1993. "Learning, Mutation, and Longrun Equilibria in Games." *Econometrica*, 61, 29–56.
- Lewis, David. 1969. *Convention*. Cambridge, Massachusetts: Harvard University Press.
- Mailath, George. 1998. "Do People Play Nash Equilibrium? Lessons from Evolutionary Game Theory." *Journal of Economic Literature*, 36, 1347–74.
- Marx, Karl, and Friedrich Engels. 1979. *Collected Works*. London: Lawrence and Wishart.
- Maynard Smith, John, and G. Dean Price. 1974. "The Theory of Games and the Evolution of Animal Conflict." *Journal of Theoretical Biology*, 47, 209–21.
- McPherson, Crawford B. 1973. *Democratic Theory: Essays in Retrieval*. Toronto, Canada: Clarendon Press.
- Polanyi, Karl. 1957 (1945). *Primitive, Archaic and Modern Economies*. London: Routledge and Kegan Paul.
- Pollock, Ethan. 2007. *Stalin and the Science Wars*. Princeton, New Jersey: Princeton University Press.
- Rousseau, Jean-Jacques. 1973 (1762). *The Social Contract*. Edited, together with the *Discourses*, by G. Cole. London: Dent.
- Schelling, Thomas. 1963. *The Strategy of Conflict*. Oxford, England: Oxford University Press.
- Sugden, Robert. 1986. *The Economics of Rights Co-operation and Welfare*. Oxford, England: Blackwell.
- . 1989. "Spontaneous Order." *Journal of Economic Perspectives*, 3, 85–97.
- . 2001. "The Evolutionary Turn in Game Theory." *Journal of Economic Methodology*, 8, 113–130.
- Varoufakis, Yanis. 2002–03. "Against Equality." *Science & Society*, 66:4, 448–72.
- Vermeij, Geerat J. 2006. *Nature: An Economic History*. Princeton, New Jersey: Princeton University Press.
- Wood, Allen. 1991. "Marx Against Morality." In P. Singer, ed., *A Companion to Ethics*. Oxford, England: Blackwell.
- Wood, Ellen M. 1989. "Rational Choice Marxism: Is the Game Worth the Candle?" *New Left Review*, 177, 41–88.