

Are we witnessing a revolution in methodology of economics? About Don Ross's recent book on microexplanation

MAURICE LAGUEUX

Université de Montréal

Abstract: The paper aims to assess whether the ideas developed by Don Ross in his recent book *Economic theory and cognitive science: microexplanation*, which relates neoclassical economics to recent developments in cognitive science, might revolutionize the methodology of economics. Since Ross challenges a conception of economics associated with what is pejoratively called "Folk psychology", the paper discusses ideas of the philosopher Daniel Dennett on which this challenge is largely based. This discussion could not avoid bearing on questions such as the nature of consciousness, the interpretation of ontological realism, the relations between agency and selfhood, and the nature and scope of economics. The paper attempts to rehabilitate the two pieces of the traditional conception of economics that were most radically contested by Ross, namely methodological individualism and the foundational role of (human) rationality in economics. A relatively nuanced judgment on Ross's bold enterprise is proposed in conclusion.

Keywords: methodology, intentional stance, consciousness, selfhood, individualism, rationality

JEL Classification: A12, B41, B52, D01, D03

It is remarkable how various contributions in methodology of economics that champion new orientations for this discipline have been published since the beginning of the new century. Recent books like *Modeling rational agents* by Nicola Giocoli (2003) and *Machine dreams* by Philip Mirowski (2002) were self-defined as contributions to the *history* of

AUTHOR'S NOTE: A first draft of this paper was presented at an EIPE Research Seminar in March 2007. For their comments, I wish to thank Menno Rol, Jack Vromen, and the other participants of the seminar, as well as to William Colish, Robert Nadeau, Christopher Pitchon, Christian Schmidt, Bernard Walliser, Don Ross himself, and two anonymous referees of the EJPE. I am also grateful to the Social Science and Humanities Research Council of Canada (Ottawa) for financial support.

economics, but they raise many questions that might have considerable repercussions for methodology of economics itself. These books were followed by two more explicitly philosophical essays, *The theory of the individual in economics*, by John Davis (2003), and *Economic theory and cognitive science: microexplanation*, by Don Ross (2005), which radically put into question some of the more respected tenets of traditional economic methodology

Many other important methodological contributions, especially papers in the main methodological and historical journals, have also been published during these few years, but my goal, here, is not to provide a survey of the work done in methodology of economics during this short period, but to inquire whether the methodological ideas proposed in such contributions can be described as revolutionary. Since it would not be possible in a short paper to seriously discuss each of them, I will focus on Ross's book, which Alex Rosenberg—who is himself one of the most respected methodologists of economics and has intensively published in the area throughout the last quarter of the 20th century—has presented as “the most important new work in the philosophy of economics in years” (jacket of Ross 2005).

Instead of developing a systematic discussion of Ross's ideas, I will emphasise what I consider illuminative and potentially “revolutionary” in them, but I will conclude that we should not exaggerate the consequences of these rich contributions nor disqualify too swiftly the more traditional approaches to methodology of economics that was developed in the last three decades of the 20th century. In order to substantiate this view, I will devote the last two parts of this paper to a discussion of the two questions which are most at risk of being affected by these new ideas, namely: methodological individualism (a favoured target of Ross), and the notion of rationality (so closely linked to the radically redefined concept of an economic agent).

FOLK PSYCHOLOGY AS A PREDICAMENT OF ECONOMICS

A trait which will be immediately noted is the fact that economic questions are integrated in much larger considerations: the developments and theoretical debates in the 20th century mathematics, the rapid progresses realised in neuropsychology, in cognitive sciences, in artificial intelligence, and in biological evolution, are considered and discussed in order to characterise what should be the proper place and status of economics. Economics here is no longer treated as a separate

science,¹ and we are witnessing, from this point of view at least, if not a revolution, at least a massive illustration of an approach that could not easily be put aside in future work in methodology. This trait is obvious in Ross's contribution, but it is his crucial theses rather than views on the scope of economics that will be mostly considered in the present paper.

To start with, one must acknowledge that Ross's book will have to be treated with circumspection, as it will be followed by a second volume, which is to focus on *macroexplanation*, and to draw on the economist Ken Binmore's contribution, just as his present volume on *microexplanation* draws on the philosopher Daniel Dennett's theses. At various places in his book,² Ross announces that he will come back to some discussed questions in his second volume, which are frequently the most typically economic ones. Nonetheless, the first volume clearly builds up the theoretical frame in which a microeconomics should be developed, and it is precisely this aspect that I would like to discuss.

To understand in what sense this frame differs so radically from the one that economists are used to, it is appropriate to recall the central problem that Alexander Rosenberg has so frequently raised throughout his various publications.³ For Rosenberg, economics suffers from a predicament that has impeded it from making the kind of progress expected of a science born more than two centuries ago. Normally, such a progress should have been observed in the accuracy of its predictions, but nothing significant has been noted from this point of view. According to Rosenberg, the predicament is that concepts like "beliefs" and "desires" (or "expectations" and "preferences" as economists prefer to say) do not "describe 'natural kinds'" characterised as "sets of items that behave in the same way, that share the same manageably small set of causes and effects". Therefore, these concepts, typical of a folk psychology, "cannot be brought together in causal generalizations that improve on our ordinary level of prediction and control of human actions, let alone attain the sort of continuing improvement characteristic of science" (Rosenberg 1994, 224). Contrary to concepts like "gene" and "acid", they have not been "carved" by a rigorous scientific analysis; they are rather inherited from a popular way of

¹ For the notion of a *separate* science as "concerned with a domain in which a small number of causal factors predominate", see Hausman 1992, 224-225, 90-97.

² For example, at Ross 2005, 291, 303, 313, 316, 320, 345, 353, 373, 381, 386, 393.

³ See, for example, Rosenberg 1988, 15-20; Rosenberg 1992, 129-131, 148-151; Rosenberg 1994, 217.

speaking, like the concept of “fish”. Such concepts are useful in business life but not in scientific analysis. A consequence of this is that the terms of theories built with them “do not correlate in a manageable way with the vocabulary of other successful scientific theories” (Rosenberg 1994, 224). Note that this does not mean that folk psychology is inefficient in prediction. On the contrary, most philosophers who aim to replace it acknowledge its astonishing efficiency,⁴ which allows us to predict with a remarkable accuracy so many decisions and actions typical of the daily life of human beings. The problem is that these predictions have not progressed like scientific ones should.

Ross and an increasing number of philosophers agree with Rosenberg’s analysis and insistently look for a way to avoid relying on folk psychology concepts judged inappropriate for scientific investigation. The conviction that this goal will be reached is nurtured by the following argument, which is repeated on various forms by all of them. There is no reason why this psychology based on folk concepts like “beliefs” and “desires” should not be progressively replaced by a scientific psychology based on neurobiological data, just like the folk astrophysics which held Ancients to believe that the sun and the whole sky turned around the Earth was totally replaced by a scientific astrophysics according to which it is rather the Earth that is moving. Similarly, a folk biology—according to which the intervention of an Intelligent Designer is necessarily required to explain the remarkable adaptation of most organisms and of specialised organs, such as eyes and hearts—was progressively replaced (while not yet in every circle) by a scientific theory based on natural selection.

Given the continuous scientific progress that has been made over the past four centuries, the folk psychology of economics should be replaced in turn. However, this kind of replacement should not be interpreted as a straightforward elimination of the so-called folk concepts in order to replace them by scientific concepts, despite the program of those who are known as “eliminativists”. Indeed, in the two paradigmatic cases evoked above, the folk concepts were designating phenomena or experiences which have not been eliminated but explained in a much more satisfactory scientific fashion. The diurnal movement of the sun and the whole sky from East to West is a phenomenon still experienced by everybody, but it was explained in a

⁴ For example, Rosenberg 1988, 15-16; Dennett 1991a, 29, 42, 43; Churchland 1984, 58-59.

much more satisfactory fashion by Copernicus and modern astronomers than by the Ancients. Similarly, the functional adaptation of living organisms is a remarkable fact that is explained in a compelling fashion by natural selection. Therefore, it would be ill advised to drastically eliminate concepts like beliefs and preferences simply because their prospect as scientific concepts is rather poor. They refer to intentional states which are constantly experienced by human beings and that could not be treated as nonexistent, even if it is judged essential that one develop an alternative and more scientific explanation of them.⁵ Such an explanation might be based on concepts related in some fashion to neurobiology, but (1) it must “save” the phenomena that we used to characterise as intentional, and (2) it must be really compelling as an explanation of their occurrence, whether as illusions, misconceptions, or whatever.

It is for having provided such compelling explanations of what was experienced that Copernican astronomy and natural selection have respectively replaced folk astronomy and folk biology. It is doubtful that the learned community would have massively rejected folk theories on these matters if, without proposing compelling alternatives to explain phenomena such as diurnal movement and functional adaptation of organisms, Copernicus and Darwin, and their respective successors, had contented themselves with claiming that, according to sound principles of dynamics, it is not reasonable to think that the whole sky turns around the Earth, or that Creationism does not fit well with the rest of biological discoveries. Put otherwise, given the irrepressible need for explanation of phenomena whose significance is perceived as major, the onus of proof lies on those who want to dislodge folk theories.

From this point of view, Paul Churchland’s attitude is somewhat irresponsible when he defends his materialist bottom up methodology with the following argument: “If the thumb-worn categories of folk psychology (belief, desire, consciousness, and so on) really do possess

⁵ John Searle (1997, 111) made an observation of this kind. He also developed a more radical argument aiming to show that the case of intentional states is crucially different from the two others I have mentioned, because “where consciousness is concerned the existence of the appearance is the reality”. He illustrates this view with the following comment: “the experience of feeling the pain is identical with the pain in a way that the experience of seeing a sunset is not identical with a sunset” (Searle 1997, 112, and the point is revisited in 121-122, 124). I think that this is an interesting point which should be considered seriously, but which can be contested by one who would claim that, even though the pain and the feeling of the pain are non separable (or even identical), the point is to determine the very nature and the origin of what is called a feeling and not to decide whether experiencing it is a real experience or not.

objective integrity, then the bottom up approach will eventually lead us back to them” (Churchland 1984, 97). How can one proceed to research by assuming that phenomena experienced by everybody like beliefs, desires, and consciousness may be forgotten and a priori reputed inexistent if they do not happen to be rediscovered by a particular type of materialist analysis?

In any case, it is the merit of Daniel Dennett, together with a few other philosophers, to have devoted his own career to the difficult task of explaining beliefs, desires and even consciousness, while constantly remaining faithful to a commitment requiring that such an explanation be derived from an analysis of the brain and of its environment. And it is the merit of Don Ross to have boldly attempted the equally difficult task of adapting such an approach to economics, a science that has almost always assumed the full validity of the traditional conception of beliefs, desires, and consciousness. One must acknowledge indeed that Dennett and Ross have emphatically rejected eliminativist theses such as Churchland’s. However, one must inquire about the ontological status granted to these intentional states, which are rescued from elimination in such a highly theoretical process.

THE ROLE OF THE INTENTIONAL STANCE

When it comes to explain why people *have* beliefs, the fundamental step in Ross’s argumentation is based on Dennett’s notion of “intentional stance”. Since the role played by the intentional stance in Dennett’s philosophy may easily be a source of confusion, it is worthwhile to recall what is involved in it. According to Dennett, it is an intellectual attitude corresponding to a strategy of interpretation that presupposes that an object (not necessarily people) has intentions and act rationally in such a way that it becomes possible for us to predict its behaviour.⁶

At first glance, the notion seems to be a very simple one, since it describes what we are so frequently doing when we say, for example, that our computer wrongly *believes* that we ask it to do an operation which, in fact, does not really interest us, or that it does not *want* to do a particular operation, or when we explain to a child that a frog *believes* that a fly is good to eat and *wants* to eat it. We may be convinced that such a parlance is metaphorical, but, for Dennett and Ross, it is not really so. Dennett admits that the metaphorical view is “immensely persuasive”, but rejects this interpretation as deceptive in favour of his

⁶ See, for example, Dennett 1987, 15; and Ross 2005, 38.

own position according to which “there is *nothing more* to *our* having beliefs and desires than our being voluminously predictable (like the frog, but more so) from the intentional stance” (Dennett 1987, 108, first emphasis added, second one is in the original). Taking an intentional stance when we consider human beings—which means granting human beings intentional states—allow us to voluminously predict their behaviour. But the behaviour of a computer or of a frog can be voluminously predicted as well when we grant them these intentional states by taking an intentional stance; therefore Dennett concludes that the difference between them and us turns out to be a simple matter of degree.

Now, the intentional stance is not the only way we have to predict the behaviour of what Dennett and Ross call a “system” (a computer, a human being, a frog, or any other animal). One may take a design stance, which consists in treating the system the way we are used to considering a machine that has been designed by an engineer. People unfamiliar with mechanics are frequently tempted to take an intentional stance toward their car which does not “want” to work properly, but, in this case, it is normally more appropriate to take a design stance and predict or explain the behaviour of the car from the examination of the functions that the engineer has reserved for the various parts of the engine. Dennett adds that a third stance which must be distinguished from these first two is the physical stance, the one we take when we predict or explain the behaviour of a physical body with the help of the causal laws of physics without assuming the intervention of any designer.⁷

Thus, in order to predict or explain the behaviour of a thermostat, one may take either a physical stance, looking at the causal laws affecting its material parts, a design stance, looking at the functional parts put into interrelation by engineers or an intentional stance looking at the thermostat as believing (rightly or wrongly) that the temperature is at such a level and wanting to avoid further heating. Dennett insists on the fact that, while the physical stance remains the most fundamental one, each of these three stances allows us to understand many phenomena that would be non-accessible from the other stances. This can be easily illustrated by an example familiar to economists. A Martian who would like to understand what is going on in an economic exchange by following, with the help of physical laws, the movements of

⁷ For a presentation of these three types of stance, see Dennett 1987, 16-18, 38-40.

commodities and of pieces of money involved in the process would totally miss the economic exchange itself. Consequently, this Martian would have a lot of trouble to predict the developments of the phenomenon, in contrast with economists who quite appropriately take an intentional stance (they grant intentional states to the traders) in order to explain it.

None of these interesting considerations sounds very offensive for traditional economics, except, maybe, for the idea that associates one's beliefs and desires with the very fact of taking an intentional stance (taken by oneself or by someone else). That does not mean that every time that one takes an intentional stance toward something, let us say a car for example, one is committed to admitting that the car actually has beliefs and desires. To avoid this misinterpretation, Dennett and Ross insistently claim that a system's beliefs and desires can be associated with an intentional stance taken toward them only when the intentional stance is the *only way* to predict or explain the behaviour of this system. This proviso about the intentional stance may sound a little odd, but the idea is that, no more than the Martian evoked above, anyone can capture what is involved in an event such as an economic exchange without taking an intentional stance. Yet, no one will deny that an economic exchange really exists; it is a *real pattern* that must be explained as anything else. If the seller and the buyer's actions cannot be explained otherwise than with the help of beliefs and desires, Dennett and Ross will say that it is precisely because these actions cannot be explained otherwise that their beliefs and desires can be said to be real.

Similarly, if the moves of a sophisticated chess-playing computer cannot be predicted otherwise than by taking an intentional stance toward it, they will say that this computer *really* has beliefs and desires. In contrast, my car has no such intentional states: the simple fact that I frequently say that it believes or wants so and so does not allow me to explain or predict anything that I cannot explain or predict much better by taking a design stance toward it or by taking an intentional stance toward myself, the driver. But let us consider this view more closely.

BETWEEN SCYLLA AND CHARYBDIS

According to Ross, the "core thesis" of intentional-stance functionalism, which is the name he gives to Dennett's philosophical approach that he equally adopts, consists of the following claim:

What it is to have intentional states—real ones, in the only sense of ‘real’ that attaches to any intentional states—is to exhibit behavioural patterns that can’t be predicted or explained without recognition of the patterns indexed by the intentional states in question (Ross 2005, 63).

As seen above, Dennett, for his part, presents the same thesis in the following terms: “there is nothing more to *our* having beliefs and desires than our being voluminously predictable [...] from the intentional stance” (Dennett 1987, 108). But what is the exact meaning of this sentence? If it is interpreted as simply saying that one’s real beliefs and desires are inferred from the fact that we can predict one’s behaviour with their help, in such a way that their real existence is confirmed, it is difficult to see in what sense this view significantly differs from folk psychology. Economists and historians who rely on folk psychology do not pretend to directly experience beliefs and desires of the agents they are studying; they assume that these agents have beliefs and desires and, since they can predict or explain their behaviour on this basis, they conclude that their assumption was well grounded. To avoid such a traditional interpretation and fully appreciate the originality of intentional-stance functionalism, one has to take seriously the words “nothing more” in Dennett’s sentence, and construe the latter as *denying that the existence of beliefs and desires means anything more than the fact that they are a necessary condition to predict a specific behaviour.*

This is the interpretation that explains why Dennett’s view has frequently been pejoratively characterised as “instrumentalist” (beliefs and desires being nothing but instruments to predict). According to Ross, one should instead characterise it as “behaviourist” (beliefs and desires being nothing but the fact that they are *required* to explain a particular behaviour), and admit that it is a consistent first step in a materialist attempt to “save” intentional phenomena from being purely eliminated. However, even if one is fully happy with this “behaviourist” way to characterise intentional states, one may raise an objection, that Ross does not consider explicitly, about the capacities of the “systems” that take such intentional stances. Who takes intentional stances? Clearly human beings take them, whether or not other systems also do. But if human beings have the capacity to take intentional stances, which means to interpret the behaviour of something by attributing it intentional states such as beliefs and desires, it is because they *already*

have beliefs and desires themselves (they desire to predict behaviour and they believe that taking an intentional stance is the proper way to satisfy this goal). Now the point of invoking intentional stances was to solve the problem raised by the questionable existence of beliefs and desires. Explaining beliefs and desires by the fact that an intentional stance was taken toward the person who has them and explaining the capacity to take this intentional stance by beliefs and desires of the intentional-stance taker clearly looks like a case of chicken-and-egg type explanation.

A possible way out of this objection might be that even if human beings take intentional stances with the help of their beliefs and desires, this does not imply that intentional stances cannot be taken in quite different ways. I will consider this view below, but first I would like to examine how Dennett, according to Ross, answered to an objection of this type, put forward mostly by John Searle, according to which “our attribution of intentional meaning to states of artifacts is parasitic on the fact that we are already intentional interpreters” (Ross 2005, 43).⁸ According to Ross, the strategy of Dennett’s answer to Searle was to explain how people can have intentional states, just like computers! (Ross 2005, 44; see Dennett 1987, chap. 8). After all, human beings have been designed by natural selection with a brain that may be described as an exceptionally versatile computer. Note that this could hardly be considered a direct answer to the question I raised, which does not directly concern the possibility for people or computers of *having* intentional *states* or not, but the fact that human beings can *take* intentional *stances* to start with.

In the chapter 8 referred to by Ross, Dennett uses a few highly ingenious mental experiments to argue that if, as easily admitted, the “intentional states” of the device of a soft drink vending machine that accepts quarters and rejects slugs are granted to it metaphorically, so is the case for an extremely sophisticated robot and for human beings (Dennett 1987, 294, see also 290-298). The idea is that, even though the device, the robot and their “states of mind” are just artefacts, people would be artefacts as well, “artefacts designed by natural selection”

⁸ With this sentence, Ross intends to capture the gist of an objection raised by Searle (in Searle 1980; see, for example 418, left column), which was implied by Searle’s defence of his famous Chinese room argument; but, in contrast with the one I proposed above, this is not an objection addressed to Dennett’s thesis on the intentional stance.

(Dennett 1987, 300). Natural selection would have designed not only our brain, which is a generally accepted view, but also the meaning which is circulating in it, just like engineers have designed not only the hardware of the robot but the software through which meaning is transmitted. Dennett does not hesitate to locate some intentionality in our genes in order to explain intentionality in our minds: “So our intentionality is derived from the intentionality of our ‘selfish’ genes! They are the Unmeant Meaners, not us!” (Dennett 1987, 298). This is granting quite a bit to natural selection and genetics; but if we accept this view, Dennett’s answer to the question I raised above might be that *natural selection* has designed human brains that have beliefs, desires, and other intentional states including the capacity to interpret and, therefore, to *take* intentional stances.

Such a view, however, would introduce another problem for intentional-stance functionalism. If such human interpreters are produced ready-made by natural selection, with intentional states and the capacity to take intentional stances, the behaviourist interpretation of Dennett’s view, according to which “there is nothing more to *our* having beliefs and desires than our being voluminously predictable [...] from the intentional stance” (Dennett 1987, 108), no longer holds. Indeed, there would be *something more* in our having beliefs and desires, namely what natural selection would have provided to us, and which was acquired without the help of any intentional stance taken toward us, since natural selection does not take intentional stances. What meaning is left of this insistent behaviourist claim that beliefs and desires are *nothing more* than the fact of being predictable if it is admitted that natural selection has made people with desires, beliefs, and the capacity to take any stance they like? Should we conclude that we are back to a “folk psychology” interpretation according to which beliefs and desires fully exist, with the additional precision that this is due to the work of natural selection?

CAN CONSCIOUSNESS BE EXPLAINED?

For his part, Ross does not rely on such an alleged genetic basis of intentionality, even though he does not hesitate to grant to human systems the capacity to take an intentional stance not only toward other systems but also toward themselves, which means to interpret reflexively their own intentional states. Thus, he maintains that “[t]he main (relevant to present issues) difference between existing chess-

playing machines and human chess-players is that the latter do, and the former don't, take the intentional stance toward *themselves*" (Ross 2005, 63). Ross, however, is fully aware of the difficulty involved in the notion of one adopting the intentional stance toward oneself. Since the subject that takes such a stance cannot "be a *part* of the system", when this reflexive intentional stance is evoked, the "most immediate and vicious sort of circularity thus seems to threaten" (Ross 2005, 286).

To avoid this threatening circle, Ross—who estimates with Dennett that deciding whether a chess machine can take an intentional stance toward itself has much to do with deciding whether it has consciousness or not—turns toward the thesis that Dennett developed in *Consciousness explained*. In this book, Dennett no longer relies on genes to explain intentionality. Instead, it is cultural selection that is invoked in order to account for human specific capacities and especially for consciousness (Dennett 1991a, 199-207). Thus, consciousness must be seen as "a product of cultural evolution that gets imparted to brain in early training" (Dennett 1991a, 219; see also Ross 2005, 160). After having rejected as a dead-end attempts to explain consciousness by looking *inside* the human brain, Dennett and Ross had little choice other than turning toward cultural and social factors to explain it (Ross 2005, 44-52); therefore, intentional-stance functionalism is a resolutely *externalist* approach.

Far from being explained by the very structure of the brain, consciousness, according to Dennett, must not be conceived as a solidly unified entity that would survey the activity of a person. It would be rather a result of multiple drafts (the word "draft" being understood here as the successive drafts of a paper written by a perfectionist author) in such a way that, each of these drafts being potentially operative at one point in time, there is no such thing as a final draft, which could be considered as "the moment of consciousness" (Dennett 1991a, 126; see also 113, 125-126) It is the reason why Dennett has named "Multiple draft model (MDM)" his model of consciousness. But how can these drafts be developed? Essentially as a result of social intercourse and by the use of public language.⁹ Now, few persons will deny the decisive role of society and of public language in the formation

⁹ "Public language" designates here a social mode of communication transmitted from generation to generation (Ross 2005, 288), which must be distinguished from an internal "language of thought" (Ross 2005, 53) and also from "emotional signalling systems" (Ross 2005, 300).

of human beings, both as a species and as individuals, but Dennett and Ross go much further. Public language is described by Ross, who makes explicit one of Dennett's suggestions, as "*the scaffold that makes humans so strikingly different in their ecology from other intelligent animals*" (Ross 2005, 286-287). Indeed, Dennett even claims that the difference between human beings, the "most prodigious intentional systems on the planet" and the poor intentional systems exemplified by frogs is largely explained by the fact that the former are "bathed in words" that allow them to "assert, deny, request, command and promise" (Dennett 1987, 112). How this bathing in words may transform us through a relatively uncontrolled use of public language is what Dennett discusses with great ingenuity in chapter 8 of *Consciousness explained*, even though one may remain unconvinced that this process can ultimately generate the capacity for human beings to take intentional stances toward themselves.

It is now possible to recapitulate. If Dennett was right in chapter 8 of *The intentional stance* when saying that intentionality is derived from our genes thanks to natural selection, we would be fully equipped to take intentional stances toward other people or even towards ourselves, but this process would be redundant since, thanks to natural selection, people would be already endowed with intentional states such as beliefs and wants. One may admire the ingenious analogies between the working of a computer and the working of the brain that Dennett explores in this chapter, but if natural selection was so generous, it is difficult to understand how one could still maintain that beliefs and desires are *nothing but* the fact that intentional stances must be taken toward those to whom such intentional states are attributed. Why should human beings—whose brains would be endowed with such a well designed program allowing them to interpret other's actions and their own actions by taking intentional stances—wait to be themselves the object of an intentional stance in order to be able to experience beliefs and desires?

In fact, Dennett and Ross do not want to go so far; while they claim, like so many thinkers, that natural selection has designed the brain as a remarkably efficient hardware, it is a *cultural* selection—based on social intercourse and public language—that they invoke to explain the development of the required software, namely intentionality and consciousness. One would like to know more about the mechanism through which such a highly sophisticated software—about which

Dennett even suggests that its survival to the brain cannot in principle be excluded (Dennett 1991a, 430, 368)—has been designed, not by a learned engineer working in artificial intelligence, not even by natural selection, but by the virtue of the progressive use of a public language learned by cultural interaction. In any case, were Dennett and Ross claiming that social intercourse and public language can really endow *people* with the capacity and the autonomy required to take such intentional stances, the “behaviourist” thesis, according to which intentional states *depend* in some way on being the object of an intentional stance, would no longer keep afloat. Indeed, here again, claiming that the joint effect of natural and cultural selection was to provide human beings with all the intentional states required to take intentional stances brings us back either to the neutralisation of this “behaviourist” thesis in favour of an eliminativist one, or to a revamped folk psychology, which could very well work with these gifts of selection.

Clearly, to understand the role reserved to the intentional stance, behaviourist and externalist components of the discussed thesis must be interpreted in a more radical fashion. As underscored by Ross (in a personal communication), taking an intentional stance “is manifested in ‘behavior’ not, in the first place, in beliefs”. This might ensure a role to intentional stances within the very process of cultural selection, but this move raises new questions. Let us admit that, before being able to believe or to want anything, it is possible to behave in a way that corresponds to taking an intentional stance, and let us try to imagine such behaviour. For example, an individual might be afraid, in a strictly behavioural manner, when facing someone else and predict an aggressive behaviour on this basis, but it is far from clear that no intentional states would be involved in such a situation, especially if we remember that Dennett showed how it is complicated to decide whether frogs really have beliefs and desires or not (Dennett 1987, 106-110). Possibly, most relevant examples of behavioural attitudes corresponding to a purely behavioural intentional stance might be found, but the difficulty becomes still more serious when we consider what is required for an intentional stance to be efficient enough to contribute in some way to the fact that desires and beliefs with those who are the objects of such stances can be considered real.

Indeed, after recalling the criterion invoked by Dennett and Ross for not being a simply metaphorical intentional stance, which is that the

intentional stance be the only possible way to predict the considered behaviour, let us suppose that a frog has no desire to avoid predators nor beliefs about the efficiency of jumping away to reach such a goal but nonetheless jumps away when facing an aggressive move of a predator. Can we really say that the frog has taken a behavioural intentional stance that may allow us to claim that the aggressive intentions of the predator were *nothing more* than the fact that they provoked this move of the frog? Answering yes might sound a bit preposterous, but if the answer is no, why would a similar move have this consequence when we replace the frog by pre-evolved human beings who are still unable to have beliefs and desires? Clearly, Dennett and Ross's answer would be that in contrast with these human beings, the frog is not "bathed in words".

But what is the exact role of public language? The point is not to simply claim that public language has strongly contributed to develop typical abilities of human beings, a claim that is perfectly acceptable even for "folk psychologists". Ross goes much further and his externalist approach may even imply that propositional attitudes such as beliefs and desires should be considered real "not as descriptions of patterns in brains, but as descriptions in patterns of social communication" (Ross 2005, 61). If we consider that the development of *real* desires and beliefs in social communication is clearly less problematic than their development in a human brain, this construal might be a way to conciliate significant intentional stances (taken from the social world endowed with desires and beliefs) with the idea that pre-evolved human beings have no intentional states. However, how such a purely behavioural intentional stance might be *socially* taken with the help of social beliefs and desires which do not exist yet in individual human beings is far from clear. And such is the idea that intentional states so generated were nothing more than their association with behaviours voluminously predicted through (and only through) this behavioural and social intentional stance.

In any case, this hypothesis, which can hardly be intuitively or empirically grounded, supposes the existence of a social agency that I will briefly discuss later. For now, let us conclude that the problem lies with the fact that the notion of individual stance, that is perfectly clear when it is attributed to fully developed human beings, becomes less and less clear when, in order to make the theory consistent, it is attributed to stance takers that have little in common with them.

INSTRUMENTALISM VERSUS REALISM

If intentional-stance functionalism seems to hesitate between endowing human beings with full capacity to take intentional stances, and saving the “behaviourist” thesis by subordinating their intentional states to a socio-behavioural intentional stance, it is because an obvious tension exists in Dennett’s thought between an “as if” *instrumentalism* associated with this reduction of intentional states to simple instruments of interpretation for those who take an intentional stance and a somewhat hesitant *realism* according to which cultural evolution has installed in human beings ontologically consistent powers.

As recalled by Ross, Dennett’s position “has regularly been associated with instrumentalism” (Ross 2005, 64, see also 160, 264). John Davis, for example, associates Dennett with Friedman and claims that, like the latter, “Dennett is not interested in the realism of our assumptions about the mind, but only about their predictive value” (Davis 2004, 96). It is true that various passages in Dennett’s work may incite readers to think this way, but the anti-instrumentalist and even realist character of Dennett’s philosophy is one of Ross’s most consistent claims. After all, what makes Dennett’s originality among the most radical of materialist thinkers is his anti-eliminativism. Against those who claim that notions like beliefs, desires, and consciousness must be eliminated, Dennett has devoted a large part of his work to show that such entities *really* exist. Ross, who claims that a basic realism is a presupposition of his own book (Ross 2005, 21-22, 57), still accentuates this realism, which for him is capital for economics, and presents Dennett, surely not as a commonsense realist, but as “radical scientific realist” (Ross 2005, 163-164).

But what about the intentional stance, which is so easily perceived as an instrumentalist trick allowing us to interpret people’s actions *as if* they were guided by desires and beliefs? To counter this perception, Ross distinguishes two quite different (while complementary) activities pertaining to intentional-stance functionalism: one is purely *methodological* (MISF), and the other is resolutely *ontological* (OISF) (Ross 2005, 336-337). It is only the former that can be said to be instrumental, for example when I attribute beliefs and desires to an object—a thermostat, for example—just because it is *useful* to predict or explain its behaviour without seriously thinking that the intentional states referred to have any ontological status. By contrast, the latter

“aims at explaining, still in intentional terms, the dynamics of systems one already has reasons for believing to be irreducibly intentional” (Ross 2005, 336).

To illustrate such a case, Ross, safely enough, takes the example of “a person”. This counterargument simply consists in accepting the validity of the accusation of instrumentalism for non-crucial cases, where the intentional stance has no special ontological pretensions, but since the accusation normally bears on the cases where intentional stances are taken toward a person with, according to Ross, an ontological meaning, introducing such a distinction is tantamount to simply rejecting the accusation in question. However, the borderline between both types of cases remains rather vague as illustrated by the case of frogs discussed non-conclusively by Dennett (Dennett 1987, 106-116, see especially 111-112). Thus, the most important question concerns the foundations of the ontological certification that is granted this way to only *some* intentional stances taken.

On what grounds does intentional-stance functionalism attribute *ontological* status to entities such as intentional states? In his remarkable paper of 1991 entitled “Real patterns”, Dennett comes back to an idea he had introduced a few years earlier in *The intentional stance* where he drew attention to the fact that entities such as centres of gravity can be said to be fully real without being pieces of “furniture of the physical world” (Dennett 1987, 72). The 1991 paper characterises such entities as *real patterns*, which are *real* because, when it comes to capturing the phenomenon to which they correspond, “*there is a description of the data [constituting them] that is more efficient than the bit map*” (Dennett 1991b, 34). For example, centres of gravity satisfy this criterion since “we think they serve in perspicuous representations of real forces” (Dennett 1991b, 29).

According to Dennett, the notion of existence should not be treated as univocal, a view which allows him to save at a relatively low cost the existence and reality of intentional states: “beliefs are best considered to be abstract objects rather like centers of gravity” (Dennett 1991b, 29). However, if it is true that the notion of “existence” can be understood as a matter of degree, one may wonder whether the totally passive and abstract notion of centre of gravity can really be put on the same (existential) footing as notions such as belief and consciousness, which correspond also to real patterns, but which have hardly any meaning at all if they are emptied from their *active* connotations.

After publishing his own paper on real pattern (Ross 1995), Ross came back to the central question raised by Dennett's paper in his contribution (Ross 2000) to an edited collection of essays on Dennett's philosophy. This contribution proposes an ontological interpretation of Dennett's views that the latter has received with an evident sympathy, which did not however imply full conviction (Dennett 2000, 359-360). Ross unequivocally claims in his chapter that "reality is composed of real patterns all the way down" (Ross 2000, 160). This must be understood as a strong ontological thesis claiming that "to be is to be a real pattern" (Ross 2000, 161), a thesis which requires a criterion for determining when a pattern can be said to be "real".

Ross formulates a technical criterion that adds a few precisions to Dennett's criterion of being "more efficient than the bit map".¹⁰ For Ross, such conditions seem necessary to protect Dennett's philosophy against a "slide into instrumentalism" (Ross 2000, 160) and to definitively put it on the side of a kind of realism that Ross has christened "Rainforest realism". It might be difficult, however, to avoid thinking that such an intricate criterion required for determining whether something is *a real being* sounds a bit ad hoc and that the way to check whether consciousness, for example, satisfies it is not very clear.

SELVES VERSUS AGENTS

Be that as it may, when it comes to account for the reality of the selves in *Consciousness explained*, Dennett describes it both as an *abstraction* (Dennett 1991a, 368, 414)—an aspect which does not fit very well with Ross's views concerning Dennett's realism—and what he calls a "*center of narrative gravity*" (Dennett 1991a, 418, 429), meaning that it is narratives and biographical accounts that "spin" a self, just as a character in a novel is spun. Ross draws from this idea concerning the formation of the self through narratives (Ross 2005, 280, 285, 286), while emphasising the role of education still more than Dennett does (Ross 2005, 282-289). Moreover, he suggests completing this analysis

¹⁰ The whole text of these conditions reads this way: "To be is to be a real pattern, and a pattern is real iff: (i) it is projectible [sic] under at least one physically possible perspective and (ii) it encodes information about at least one structure of events or entities S where that encoding is more efficient, in information-theoretic terms, than the bit-map encoding of S, and where for at least one of the physically possible perspectives under which the pattern is projectible, there exists an aspect of S that cannot be tracked unless the encoding is recovered from the perspective in question" (Ross 2005, 68-69; but also see Ross 2000, 161).

with the help of game theory by distinguishing three levels of games that would be played in the process of the acquisition of the self. However, probably more interesting for the methodology of economics than these levels—whose relations are carefully described, but which are not directly related to concerns in economics, and are not concretely implemented—is the idea that Ross develops about the relation between agency and selfhood.

In *Consciousness explained*, Dennett reported scientific observations of multiple personality disorders, to the effect that “a single human body seems to be shared by several selves” (Dennett 1991a, 419ff.). Ross made some headway on observations of this kind which allowed him to conclude “the Dennettian theory that separates agency from selfhood conceptually undermines microeconomic individualism” (Ross 2005, 311). Now, Dennett is not an economist and he rarely refers to economics and even relatively rarely to the notion of agent (or agency). In the indices of these books, “agent” appears only once in *Consciousness explained*, nowhere in *The intentional stance*, and ten times in *Brainstorms*; but none of these uses is related to economics or to the concept of self, which never appears in the same chapter as the concept of agent. For Ross, who is an economist as well as a philosopher, the notion of agency was a central one, and the idea that it could not be hardwired to the notion of self, or to the notion of human being, became a central point of his analysis of economics.

This idea plays a decisive role in his discussion of Gary Becker (even if it is only a marginal piece in his book), and above all in his discussion of the economics derived from Samuelsonian revealed preference theory (which is a central piece in the economic application of his theory). Since he does not see any problem in dissociating agents from selves, Ross can claim that “the biography [which is related to the self] of a typical person can’t be the biography of a single (diachronic) economic agent” (Ross 2005, 156). According to him, only a contestable Aristotelian assumption incites us to think that human lives should be modelled “as single projects aimed at achieving (some) consistent goals” (Ross 2005, 159). Thus, if there are such things as human agents (and utility functions), they do not have to be coextensive with selves (and their biographies).

Ross’s position in this discussion was based on his decisive adoption of revealed preference theory (RPT) as the paradigm of neoclassical theory. Many economists would contest such an assessment, and their

contestation would probably not be diminished by the fact that this assessment is based on philosophical as much as economical considerations. Indeed, Ross took “Samuelson’s generic late-positivist philosophy more literally than he did” (Ross 2005, 156). He illustrates this by admitting that its usual application to typically economic matters was not relevant in his considerations: “I have treated RPT just as a set of axioms, leaving completely open the question of which phenomena, if any, the axioms describe” (Ross 2005, 156). A consequence of this decision is that Ross’s conception of an agent is strictly determined by these axioms as illustrated by the discussion of Becker’s thesis on stable preferences: “Of course, as a matter of logic, an *economic agent* must have stable preferences; otherwise RPT does not apply to it” (Ross 2005, 157). Indeed, how can transitivity be conciliated with changing tastes?

But this decision has still more radical consequences. Once it is admitted that any straightforward agents must strictly respect the Samuelsonian axioms, which imply consistency, it is clear that human beings are disqualified, given their lack of consistency clearly manifested in experimental economics to which Ross devotes a well-documented section of his book.¹¹ But where can we find agents if human being are such poor candidates for this title? For Ross, there is a crowd of other candidates, and among them we find various inferior animals and especially insects, whose paradigmatic quality of agent he so frequently refers to: “[a] good example of a prototypical economic agent is an insect”.¹² This is hardly surprising, since insects have little propensity to modify their behaviour and, consequently, to fail a consistency test. Moreover, they are good optimisers as, incidentally, Marx had noted in a famous passage of *Das Kapital* according to which bees are able to surpass many architects in the construction of her cells, even though, the worst architect is still superior to the best bee for being able to build a house in his or her head before building it in the real world.¹³

For Ross, however, the criterion is rather “the central locus of control” (Ross 2005, 381), which is present in bugs, but not in humans. It is not clear on what ground one can rely on this central control to generate consistency and maximisation, but it seems that insects are

¹¹ Ross 2005, 165-190; see especially pages on preference reversals that experimentation has put into light (Ross 2005, 177ff.).

¹² Ross 2005, 251; but see also 95, 241, 252, 253, 256, 290, 331, 377, 381, 393.

¹³ *Das Kapital*, Book one, Third section, chapter VII, part I.

successful from this point of view. In any case, there is no doubt that some electronic devices can be modelled with the capacity to satisfy the criterion, but a bit more unexpectedly, the other kind of candidates that are systematically considered by Ross are neurons. Paul Glimcher has developed a research program called neuroeconomics, consisting of the application of economic analysis to neural behaviour. According to Ross, “the economics in question must obviously be Samuelsonian, since no one imagines that parts of organisms have utility functions based on internally represented preferences” (Ross 2005, 325). Thus, brains would be “an ideal site for Samuelsonian microeconomics” where the agents are neurons and perhaps modules (organised groups of neurons) (Ross 2005, 334). About one of these modules, the visuomotor cortex, Ross even concludes that “it is thus a straightforward economic agent, just like an insect” (Ross 2005, 331).

Even if these modules were straightforward sub-personal agents, taking an intentional stance toward them would be purely methodological; it helps to understand their working, as Ross admits, without implying that they really have intentional states. By contrast, according to Ross, “the various long-, medium- or short-term interests” (Ross 2005, 337), which were analysed, equally with economic tools, by the psychiatrist George Ainslie in his *Picoeconomics*, are such that an *ontological* intentional stance can legitimately be taken toward them, which means that their dynamics is considered “to be irreducibly intentional” (Ross 2005, 336). So much so that each interest is considered to be “as clever as a person” in such a way that “their strategic cunning will tend to unravel all equilibria” (Ross 2005, 345), at least in some types of games.

Now, the idea that human beings are constantly divided between trends that draw their decisions in various directions is an old one, which goes back at least to Augustine, but, according to Ross, this tension is explained by the fact that a typical human is constituted by a “colony of agents” which “emerge under analysis as a complex assembly of buglike homunculi” (Ross 2005, 252). This leads us to what Ross designates as “this book’s central thesis”. It might look odd that this central thesis does not directly concern economics—at least as it is traditionally understood—but rather the very nature of a person, which is defined as “a set of basically compatible long-range interests that have co-opted a sufficient army of short-range interests into their coalition to maintain stable equilibrium” (Ross 2005, 351), but one must

recall that the book bears on microexplanations in both economic theory *and cognitive science*.

A consequence of this is that “selves aren’t straightforward economic agents. They are more like nations than insects” (Ross 2005, 290). However, since people do take actions, they should be agents in some sense; according to Ross, they are “agents-by-extension” (Ross 2005, 256), in contrast with straightforward agents. Therefore, they are put on the same footing as nations, which can take actions just as well.¹⁴ Consequently, “[t]he application of economics to people will thus have to follow the same methods, and meet the same ontological and epistemological demands, as the application of economics to countries and corporations” (Ross 2005, 257). Since they are no more unified than nations, it is not surprising that, when treated as agents, humans “show ubiquitous preference reversal and time inconsistencies” (Ross 2005, 253). Instead of acknowledging that human agents, who used to be prototypical agents, are far from being consistent, Ross defines agency through consistency and concludes that humans are not prototypical agents.

The idea that human beings could be constituted of many centres of decisions is not new. As underscored by Ross, Davis, who has devoted his own 2003 book to this question, has pointed out that it must go back at least to Hume. More recently, Jon Elster edited a book, entitled *The multiple selves*, gathering papers that gave a fair idea of the state of the question in mid-1980s.¹⁵ It is interesting to note, however, that Ross bases his intrapersonal community thesis on the idea that the person, whether or not identified with a self, is a *community of many agents*, whereas most of his predecessors refer, more or less metaphorically, to a *multiplicity of selves* who are present in a single person, usually identified with an agent. Davis, for example, suggests that individuals of neoclassical economics are constituted as a community of selves; but since he is looking for “an adequate conception of the individual in economics” (Davis 2003, 80), he turns away from orthodoxy and adopts

¹⁴ Ross does not seem to be bothered by the fact that nations are intermingled between themselves, vaguely circumscribed and constantly redefined according to people’s sensibilities and ideologies.

¹⁵ Elster 1987. However, in his Introduction to the book, Elster insists on the fact that most of these tensions inside a person do not imply the duplication of selves (Elster 1987, 10, 13, 14, 23, 24, 26-27, and especially 30); however, there are two mentions of a possible exception to this, associated to the phenomenon of self-deception, on pages 28, 31. Davis also, in note 9 of his chapter 4 (Davis 2003, 196) quotes Elster, Steedman and Krause, as well as Kavka, who all deny to be claiming that there are *literally* multiselves.

the “socially embedded individual conception” (Davis 2003, 117) developed in heterodox economics. In contrast, Ross, who does not care for the unity of the individual, turns toward a version of neoclassical economics based on Samuelson’s revealed preferences, which Davis denounces for its formalist indifference to the question of individuals (Davis 2003, 93-94).

CAN METHODOLOGICAL INDIVIDUALISM SURVIVE?

These circumstantial remarks are far from doing justice to the careful analyses of Davis; and even Ross’s book, whose discussion was the core of the present paper, is so dense that only some of its main theses have been discussed. However, these books and a few others recently published force us to reconsider some of the fundamental pieces of traditional methodology of economics. These books ruthlessly reject methodological individualism and they invite us, at the very least, to reconsider the relation between rationality and agency.

What can still be said in favour of methodological individualism, which was the favourite target of so many theoreticians? Personally, I always had trouble to see what exactly was the point of the long debate between methodological individualism and the opposite thesis, whether holism, collectivism, or whatever. Of course, straw men built up by opponents to define either holism or individualism correspond to strongly opposite perspectives, but it is very difficult to find serious defenders of these extremist theses. Usually, those who consider themselves as champions of one of these theses, in order to make their position tenable, take care to introduce so many nuances in their characterisation of their favourite “ism” that the division between such opposing views becomes more or less blurred.¹⁶

On the one hand, it is not difficult, for example, to underscore various individualist features in the methodology of a radical collectivist thinker such as Marx and of a *macroeconomist* such as Keynes.¹⁷ On the other hand, it should be admitted that methodological individualism cannot be separated from the mechanism through which unwanted consequences of human actions—consequences that explain most economic phenomena like an endogenous increase in the price of

¹⁶ Malcolm Rutherford argues for a middle way between extreme versions of holism and individualism, which are “taken to be unacceptable” and “unappealing” (Rutherford 1994, 50; see also 36-37).

¹⁷ For Marx, see Elster 1985; Lagueux 2001, 698-701. For Keynes, see Lagueux 2001, 696-698.

potatoes or a rise in the rate of unemployment—result from the interference of a multitude of anonymous but, in principle, understandable individual actions.¹⁸ Very few methodological individualists, if any, would deny that the consequences of these multiple (rational) actions are constantly modified and deviated by the unforeseeable impact of the natural and, above all, social environment.¹⁹

It is also important to reject any association of methodological individualism with the imaginary economics of Robinson Crusoe, since methodological individualism is a methodology adapted to essentially *social* sciences such as economics. As aptly underscored by Ross himself (Ross 2005, 216ff.), Robinson Crusoe is just a useful pedagogical device, but methodological individualism has nothing to do with this kind of pedagogy. More importantly, methodological individualism is *not* a reductionism. The point of methodological individualism is not to *reduce* social phenomena to individual ones; given what was said above about unwanted consequences, such a reduction would be doomed to fail. It is to *understand* social phenomena by explaining why rational human actions produce social consequences significantly different from those which would be expected by people who take such actions.

Now, human actions have a lot of social consequences, but it is clear, as abundantly illustrated by Ross, that the causal link goes at least as much the other way around. Society influences individuals possibly still more than individuals influence society. No doubt that cultural evolution of humanity can hardly be understood otherwise than as a complex interaction between societies and individuals. However, most phenomena that are explained by interventions of society have traditionally been excluded from the domain of economics. Ross has shown how far social structures, through public language, education and imitation have been determinant for the very genesis of individuals, but the genesis of individuals concerns traditionally anthropology,

¹⁸ Among others, J. W. N. Watkins, who was among the first authors to devote important papers to methodological individualism, underscored the link between this methodology and unwanted consequences (Watkins 1953, 26), a notion so closely associated with the thought of Friedrich Hayek, another defender of methodological individualism.

¹⁹ In a recent paper, Geoffrey Hodgson (Hodgson 2007, 220, 222) claims that methodological individualism understood in this fashion is misnamed since an individualism that makes room for explanations requiring interaction between people is not a pure individualism, and therefore cannot easily be distinguished from approaches that are not considered individualistic. He is surely right, but my point concerns the validity of the methodology not of the label, since it is the former and not the latter that is challenged by Ross's arguments.

psychology, and neurobiology, but not economics which is concerned by the genesis of a subset of *social* phenomena. And methodological individualism is relevant only when it comes to explain social phenomena.

Ross could oppose at least two considerations to this position. Since, according to him, people are assemblies of homunculi (whether, neurones, modules, or interests), the relation between people and societies is mirrored by the relation between such homunculi and individual people; a state of things which suggests that microeconomics should analyse the relation between homunculi and individual people as well as the relation between people and societies. However, a well-grounded analysis of the way these sub-personal agents are more or less coordinated and related to the whole person might be a great triumph for behavioural psychology and for neurobiology, but not for economics as such, even if neurobiologists use RPT or other economic tools. Here it comes, the second and more important of Ross's objections, which radically rejects the traditional distinction between the respective domain of economics and of psychological and cognitive sciences.

Ross even presents his conception as based on what is probably the most respected tenet of the methodology of economics, namely Robbins's famous definition of economics as "the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses" (Robbins 1935, 16; quoted by Ross 2005, 87). Three lines further, Ross adds, however, that he drops from it the word "human", a move which allows him to enlarge the scope of economic analysis in such a way that economics becomes the science (with Robbins's provisos concerning ends and scarce means) of the behaviour of insects (and other animals), robots, neurons and interests as well. Those whom Ross pejoratively refers to as "humanists" are apparently those who would object to dropping this quite significant element of Robbins's definition.²⁰

Be that as it may, Ross can defend his position by arguing: (1) that Robbins's definition encourages a generalised and formal conception of economics rather than a conception restricted to the questions related to material wealth; (2) that Samuelson has already developed a theory which is neutral from the point of view of the object to which it is applied; and (3) that game theory and a few other techniques have

²⁰ Davis and Mirowski count among those who Ross considers humanists (Ross 2005, 46, 70, 118, 257, 258, 270).

broken the frontiers between the various sectors of the larger science concerning any kind of behaviour. For sure, such a general science of behaviour cannot be monopolized by economics, but economics can offer some useful tools to it in such a way that an economist may be tempted to intervene qua economist in this larger domain.

If such was the case, it is clear that methodological individualism would not be a methodology adapted to *this* general science, except in its restricted domain specifically concerning the social consequences of agents' actions. However, if it can still be appropriate to invoke methodological individualism (as described above namely as a strategy for understanding the social consequences of human actions) when dealing with *human* microeconomics, it is not because the latter benefits from an indefensible epistemological privilege, but, because, human beings—the only “systems” who have to understand something—*need* to understand why the consequences of *their individual actions* are typically incorporated in social structures that escape them and seem to impose their laws on them. And this is precisely the type of explanation that a methodologically individualistic economics can offer them.

RATIONALITY AND AGENCY

Another aspect of traditional economics that has to be reconsidered is the fundamental principle of rationality, which is necessarily linked to the notion of agency, since it is decisions and actions that are labelled rational or irrational. Throughout the second half of the 20th century, economic rationality was progressively associated with consistency, an association that characterises what Giocoli (Giocoli 2003) has called a “system of relations” in contrast with a “system of forces” in which rationality was rather associated with the notion of maximisation. As is well known, rationality-consistency is closely related to RPT, which taken together constitute the core of the neoclassical economics that Ross defends and that Davis criticises. RPT and rationality-consistency require a type of agency that leads Ross to declare that insects and neurons—but not humans—are straightforward agents, and that brings Davis to look in heterodox economics for a more satisfactory notion of an individual agent.

I suggest that the source of such opposite reactions to rationality-consistency lies chiefly in the way the notion of rationality has evolved with economic theory. A certain concept of rationality was already playing a central role in classical economics, especially when the first

theories of the market were developed. However, the rationality involved was a relatively *minimal* rationality,²¹ requiring just what is necessary to induce a buyer to stop paying more when the possibility to pay less for the same service becomes clearly available, or to incite a farmer to produce more wheat when the price obtained for a barrel is largely superior to the cost involved and to lower production when the contrary situation prevails. I call this type of rationality minimal because it just requires that people not be so thoughtless to produce more and more units when it is clear that they lose a lot of money every time they produce a new unit, which is roughly similar to requiring someone be only so mindful to turn off the faucet when the bathtub is full! Happily most people have at least this degree of rationality that is enough to allow a market to work more or less properly.

With the so-called marginalist revolution, economists have introduced a more precise notion of rationality. If people are not that thoughtless, why not strictly maximise whatever valuable they can obtain? This seemed to be the logical way to elaborate, with the help of calculus, a much more precise economic analysis. But treating economic agents as utility maximisers relied on a questionable psychology. As it is well known, the next eight decades or so were largely devoted to de-psychologise economic theory, a process that has culminated with RPT.

There is little doubt that this long process going roughly from Jevons to Samuelson's respective contributions corresponds to a tremendous theoretical, if not empirical, progress. There is little doubt that, thanks to the analyses developed during this process, various economic phenomena were literally discovered and others were understood with much greater precision. But this was realised through a set of systematic idealisations of the notions of rationality and the corresponding notion of agency that required forgetting the specific features of the minimal rationality that guides actual human agents.

Human beings rarely maximise and are far from being consistent; they hesitate, make mistakes, change their mind, regret, suffer from myopia, are sensitive to frames in which questions are raised and are

²¹ This notion was introduced by Cherniak (1992), but I use it in a slightly different way and context. More generally, the idea of dissociating rationality from maximisation was explored by different economists, the most influential being Herbert Simon with his notion of "bounded rationality". However, maximisation and consistency remain by far the conceptions of rationality that economists evoke the most spontaneously. In any case, in the present context, I refer to minimal rationality uniquely in order to question the way in which Ross derives so many conclusions from his adoption of RPT and strict consistency as the criterion of rationality.

influenced by superficial similarities, but they are not stupid for all that, and they make decisions and take actions in order to reach their ends. This is exactly the kind of rationality that the theories that *explain* economic phenomena require. When they invoke nothing more than this minimal rationality, these theories are protected from the otherwise devastating objections such as those that are raised by experimental economics, whose conclusions are generally that people rarely maximise and are inconsistent without being thoughtless for all that.²² Idealisations of economics are very helpful, but if we chose one of them—possibly the most distant from the actual behaviour of human beings—as the prototype of rationality and agency, it is clear that human agents will be disqualified as straightforward agents, a status which will be, by hypothesis, reserved to “systems” who satisfy the assumptions of the theory chosen.

It is great that Samuelson’s revealed preference theory can be applied to entomology, artificial intelligence, and neurobiology, in spite of being non applicable where *human* behaviour is concerned; however, for those who are specifically interested in the phenomena which are covered by the “humanist” notion of economics, the Samuelsonian agent remains a theoretical construction useful only to illustrate the working of an agent that is idealised to the point of being able to compete with an insect in matter of perfect consistency (or perfect rigidity)!

It seems reasonable to conclude that human economic agents perform actions after being involved in the more or less complex and erratic reflective activity that their consciousness makes possible and that, while they choose certain means judged to be optimal, they rarely aspire to strictly maximise the way bugs can do at their own level. This is in no way denigrating the program of research promoted by Ross, which aims to shed some light on the way the highly complex resources of the brain might help to account for the behaviour of economic agents, straightforward or not. But the question is whether this valuable program of research should be substituted for the more traditional economic program in such a way that the whole methodology of economics, including the very definition of economics, be radically transformed. This brings us back to the question that opened the present paper.

²² For an attempt to justify such a claim, see Lagueux 2004.

A REVOLUTION OR NOT?

Are we witnessing a revolution in methodology of economics? Note that I found appropriate to raise the question this way, even though neither Ross nor the other authors that I have mentioned at the beginning of this paper have explicitly pretended to promote a revolutionary way to understand economics. In any case, I think that the publication of their respective books testifies that methodologists of economics can no longer simply ignore the questions raised by the integration of this discipline in the context of the rapid development of cognitive sciences and of artificial intelligence.

It is not clear, however, to what extent these developments may have transformed the way economists should treat the questions which were traditionally their own. It is true that the rate of innovation in the methods used by economics is relatively high in the present times, especially with the increasing place occupied by game theory, but, as far as I know, it was mostly classical game theory that was involved in the spectacular transformation of conventional economics in the last two or three decades.

Ross refers more systematically to evolutionary game theory, which can find various applications in the whole domain corresponding to his larger conception of economics, but the impact of this kind of game theory on the questions traditionally treated by economists is less manifest: at least, this impact was not made clear by Ross in the present book, which however must be completed by a forthcoming second volume, more promising on this ground. Indeed, its proper subject-matter, macroeconomics must be reinterpreted with the help of evolutionary game theory in a way that might propose an original solution to the persistent problem of micro-macro relations.

But, for the time being, the proper question to ask is whether microeconomics has been endowed with the type of categories that Rosenberg hoped to see in this discipline. Folk categories such as beliefs, desires, and consciousness may have been tentatively explained with the help of combined neurological and socio-anthropological analyses, but one may wonder in what sense their alleged dependency on intentional stances constitutes a gain from this point of view. The paradigmatic example proposed by Rosenberg was the folk notion of a fish that was replaced by anatomical concepts that “cut nature at the joints” because they are defined on the basis of a genetic analysis in a

way that makes them perfectly precise and apt to make accurate predictions possible.

The concept of a (straightforward) agent has been precisely redefined by Ross along the lines proposed by Samuelson, but its direct applications do not concern traditionally defined economic questions. In contrast, the notions of beliefs, desires, consciousness, and selves, as redefined by Dennett and Ross, can hardly be safely described as cutting nature at the joints. It does not seem much more appropriate to refer to a new explanation of economic phenomena which should replace folk explanation like the Copernican scientific explanation of diurnal movement replaced the one provided by folk astronomy. Irrespective of the appreciation we may have of the validity of Ross's new type of explanation, a notable difference concerns the fact that, in this case, it is the explanative concepts of folk economics (beliefs and desires) which are themselves explained otherwise, and not the traditional explanation of economic phenomena (markets, level of prices, cost and production, and so on) which is replaced by a more scientific one, at least in the present book.

May we say, at least, that a serious attempt has been made to replace with scientific and empirical foundations the philosophical bases of traditional concepts used by economists? At some points in his book,²³ Ross insists on the importance of granting primacy to empirical and scientific considerations over philosophical ones. This seems to me a very sound principle to follow when both of these considerations really point in rival directions. If you are interested in the causes of the 1929 economic crisis, I strongly recommend you to turn toward empirical analyses provided by economists or historians specialised in this question rather than toward philosophical speculations about the origin of crises. But such situations are rather rare.

Happily, most philosophers have learned to avoid attempts to offer answers to questions reserved to specialists of an empirical domain. But when the most fundamental questions are involved, the debate is not between philosophical and empirical considerations, but between differently oriented philosophies. Sometimes, the speculations of one of them are more anchored in recent scientific developments, but that does not warrant them superiority as philosophical speculations. The history of philosophy is full of cases of unfortunate philosophical interpretation of various up-to-date scientific theories. The most famous of them, the

²³ For example, Ross 2005, 124.

positivism of Auguste Comte, was conceived by his author as the first philosophy to be really based on scientific and empirical considerations, but it opens the road to series of divagations leading to the proclamation of the dogmas of the religion of Humanity. Don Ross's philosophy is far from being threatened by this kind of divagations; however, it remains a philosophy that intelligently takes account of an impressive quantity of empirical results, but which, for an important part, is highly speculative and controversial. In spite of the fact that it may open new roads for possible inquiries that can put us on the track to potential revolutionary developments of great interest for economics, one cannot conclude that this controversial philosophical contribution, in its present state, should lead economists to massively redefine their concepts, or even to follow its author along the way out of traditional economics, which has been opened by Samuelsonian revealed preference theory.

REFERENCES

- Cherniak, Christopher. 1992 [1986]. *Minimal rationality*. Cambridge (MA): MIT Press.
- Churchland, Paul. 1984. *Matter and consciousness: a contemporary introduction to the philosophy of mind*. Cambridge (MA): MIT Press.
- Davis, John. 2003. *The theory of the individual in economics*. London: Routledge.
- Dennett, Daniel. 1981. *Brainstorms*. Cambridge (MA): MIT Press.
- Dennett, Daniel. 1991a. *Consciousness explained*. Boston: Little, Brown, and Co.
- Dennett, Daniel. 1991b. Real patterns. *The Journal of Philosophy*, 88 (1): 27-51.
- Dennett, Daniel. 1993 [1987]. *The intentional stance*. Cambridge (MA): MIT Press.
- Dennett, Daniel. 2000. With a little help from my friends. In *Dennett's philosophy: a comprehensive assessment*, eds. Don Ross, Andrew Brook, and David Thompson. Cambridge (MA): MIT Press, 327-388.
- Elster, Jon. 1985. *Making sense of Marx*. Cambridge: Cambridge University Press.
- Elster, Jon. 1987 [1985]. Introduction. In *The multiple self*, ed. Jon Elster. Cambridge: Cambridge University Press, 1-34.
- Giocoli, Nicola. 2003. *Modeling rational agents*. Cheltenham: Edward Elgar.
- Hausman, Daniel. 1992. *The inexact and separate science of economics*. Cambridge: Cambridge University Press.
- Hodgson, Geoffrey. 2007. Meanings of methodological individualism. *Journal of Economic Methodology*, 14 (2): 211-226.
- Lagueux, Maurice. 2001. Individualisme, subjectivisme et mécanismes économiques. *Dialogue*, 40 (4): 691-722.
- Lagueux, Maurice. 2004. The forgotten role of the rationality principle in economics. *Journal of Economic Methodology*, 11 (1): 31-51.
- Mirowski, Philip. 2002. *Machine dreams: economics becomes a cyborg science*. Cambridge: Cambridge University Press.
- Robbins, Lionel, 1935 [1932]. *An essay on the nature and significance of economic science*. London: Macmillan.

- Rosenberg, Alexander. 1988. *Philosophy of social science*. Boulder, Colorado: Westview Press.
- Rosenberg, Alexander. 1992. *Economics: mathematical politics or science of diminishing returns?* Chicago: The University of Chicago Press.
- Rosenberg, Alexander. 1994. What is the cognitive status of economic theory? In *New directions in economic methodology*, ed. Roger Backhouse. London: Routledge, 216-235.
- Ross, Don. 1995. Real patterns and the ontological foundations of microeconomics. *Economics and Philosophy*, 11 (1): 113-136.
- Ross, Don. 2000. Rainforest realism: a Dennettian theory of existence. In *Dennett's philosophy: a comprehensive assessment*, eds. Don Ross, Andrew Brook, and David Thompson. Cambridge (MA): MIT Press, 147-168.
- Ross, Don. 2005. *Economic theory and cognitive science: microexplanation*. Cambridge (MA): MIT Press.
- Rutherford, Malcolm. 1994. *Institutions in economics: the old and the new institutionalism*. Cambridge: Cambridge University Press.
- Searle, John. 1980. Mind, brains and programs. *The Behavioral and Brain Sciences*, 3: 417-457 [Including Searle's introductory paper: 417-424), Open peer commentary: 424-450, and Searle's response: 450-456].
- Searle, John. 1997. *The mystery of consciousness*. New York: New York Review Books.
- Watkins, John William Neville. 1953. Ideal types and historical explanation. *British Journal for the Philosophy of Science*, 3 (9): 22-43.

Maurice Lagueux is currently associate professor in the Department of philosophy at Université de Montréal where he has also taught history of economic thought in the Department of economics. His research interests are economic methodology, philosophy of architecture, and philosophy of history. Contact e-mail: <maurice.lagueux@umontreal.ca>