

Review Article

Ross, Don. *Economic theory and cognitive science: Microexplanation*. Cambridge, Mass.: MIT Press, 2005. ISBN 0 262 18246 7.

Setting the scientific cat among the humanist pigeons¹

Andries Gouws

School of Philosophy and Ethics
University of KwaZulu-Natal
Durban
4041
South Africa
Email: gouwsa@ukzn.ac.za

Abstract

This is a review article of Ross (2005), a book which attempts to show the implications of cognitive science and economics for each other. Ross makes neoclassical economics central to the unification of the behavioural sciences, and defends its fundamental health against its critics. He locates the source of the empirical and conceptual problems besetting neoclassical economics in the mistaken assumption that the economic agents neoclassicism talks about refer directly to real, whole people. Ross argues that people are *atypical* as economic agents; bugs and other non-social animals are much more typical, and neoclassical economics applies to them in a straightforward way. How then does economics apply to the atypical agents that people are? Ross's answer builds on Ainslie (2001), who models people as being the result of strategic interactions between successive economic agents, and on Glimcher's (2003) work on the neuroeconomics of game playing in monkeys which, Ross argues, can be extrapolated to humans. The author is impressed by the boldness of Ross's vision, and the arguments offered to support it. However, doubts remain about Ross's theory of the self, the lifespan of the proposed successive economic agents out of which real people are composed, the implicit suggestion that in economics approximations can be overcome, and finally the conclusiveness of Ross's case against the need for revolutionary re-appraisals of neoclassical economics.

¹ My thanks to Julia Clare, John Collier, Deepak Mistrey, David Spurrett and other colleagues in the School of Philosophy and Ethics at the University of KwaZulu-Natal, and John Hart from the School of Economics and Finance of the same university, for invaluable comments on previous versions of this article. My discussions with John Hart made me doubly aware of the enormous gaps in my layperson's knowledge of economics and its history – gaps whose effects I have doubtless only marginally remedied since talking to him.

In his ambitious book *Economic theory and cognitive science* (henceforth: *ETCS*),² the first of two projected volumes, Don Ross situates economics vis à vis the rest of cognitive and behavioural science. He wants to show the implications of new work in cognitive science for economics, and simultaneously that economic behaviour is an important testing ground for theories in cognitive science. This first volume is about the relation between microeconomics and the theory of agency. (The 2nd volume will be on macro-explanation). In tackling the theoretical problem of agency, Ross builds on the (Dennettian) theory of intentional behaviour. Ross thinks that being clear about agency will offer the key to solving various longstanding methodological and theoretical issues in economics.

A crucial aspect of Ross's project is his attempt to boil neoclassic economic theory down to its essentials. His aim is to give economic theory the maximum scope possible, while not going beyond the minimum commitments needed for embracing its formal apparatus. This involves identifying, and weeding out, non-minimal, unnecessary and counterproductive assumptions that others have made about this apparatus. It also involves a break with folk-psychological and folk-economic construals of key terms in economics, which to Ross lie at the root of many problems in the philosophy of economics.

Ross repeatedly situates his position vis à vis that of two foils (14) who are both critical of the meeting of economics and cognitive science that he pleads for. The first is Mirowski's (2002) postmodern account of the history of economic theory since WWII, which is ironically deflationary of its ambitions of becoming a 'cyborg science' – an account to which Ross is, apart from this deflation, largely sympathetic. The second is Dupré's (1993: 2001) 'humanist' attack on the ideal of a unified behavioural science.³ This strategy continues right up to the final chapter, where Ross summarises his own position by way of a reasoned overview of his differences with Mirowski and Dupré.

Dupré's alarm at the partnership between economics and cognitive science derives from a view of these disciplines that Ross challenges on every point (383). If we accept Ross's ontological and epistemological critique of Dupré,⁴ the latter's *normative* pronouncements condemning the prospect of a systematic science based on the meeting of economics and cognitive science are unsupported. What Dupré condemns as *imperialism*, Ross celebrates as *cosmopolitanism* (27-28).

According to Dupré (1993: 2001) the idea of unity of science presupposes that a wide-ranging mereological reductionism is possible *in general*. Ross (318) agrees with Dupré that such a reduction is not possible – chemistry can for instance not be reduced to physics, and there will be cases in which intentionality is irreducible (334). Notably, Ross thinks that macroeconomics will not even in principle be reducible to microeconomics. (Whether such reduction is possible, has been the topic of protracted philosophical and methodological debates regarding economics). However Ross denies that the unity of science depends on the possibility, across the board, of mereological reductionism. We could thus summarise Ross's position as 'unity of sci-

2 To avoid constant repetitions of (Ross 2005: *p. nn*) in this article, I refer to pages in Ross's book simply by numbers in brackets.

3 Davis is a third interlocutor who is regularly addressed. His book *The theory of the individual in economics* (Davis 2003) ranges over many of the same issues as *ETCS*.

4 Though there is important overlap between Wittgenstein's philosophical psychology and Dennett's intentional stance – a mainstay of Ross's book, as we will see – Ross's critique of Dupré by extension applies to much of what has been said about the social sciences in the name of Wittgenstein, and doubtless also by Wittgenstein himself.

ence without reductionism'. His own argument for the unity of science relies, rather, on *informational* notions (66-67; compare Ladyman et al. 2007).

A few years ago I would have found a project of the sort Ross engages in, unconvincing and of no relevance to the areas that interest me, such as psychoanalysis and aesthetics. That I am enthusiastic about Ross's book thus testifies to a sea change in my views. How can a work like this bear on psychoanalysis? (Space does not permit a discussion of its bearing on aesthetics).

One of the mainstays of Ross's argument is the work of George Ainslie, the American psychiatrist and behavioural economist. I was led to Ross's book via my prior interest in Ainslie's work. Ainslie's (1992, 2001) theory is exemplary for the meeting of economics with the other behavioural sciences that Ross heralds. The initial interest of Ainslie's theory for me was that it proposes a new metapsychology for psychoanalysis. This metapsychology again takes seriously the quantitative model – the quantitative model of motivation – that Freud often explicitly espoused (for example in his *Project for a scientific psychology* (Freud (1975/1895); Gouws and Cilliers 2001), and that elsewhere could seem to be no more than a metaphor; a quantitative model that many commentators saw as an embarrassing nineteenth century mechanistic holdover, to be overcome by a fully hermeneutically conceived psychoanalysis.

Freud didn't go all the way to a mathematical formulation of the metapsychology, just as Bentham, the father of utilitarianism, didn't go all the way to turning his quantitative intuitions into a mathematical formulation of the principles of the utilitarian calculus. Ross convincingly argues in a previous work (Ross 1999) that neoclassical economics is the descendant, and explicit mathematical elaboration, of the fundamental quantitative intuitions found in Bentham. Freud's thought had deep roots in utilitarianism – he was for instance the German translator of a number of John Stuart Mill's works, and the many parallels between his thought and that of Bentham have been pointed out by Watson (1958).

The convergence between behavioural economics and behavioural psychology, which has by and large vindicated the basic behavioural principles formulated by Skinner, is widely recognised today. Ainslie works at the site of this convergence, and claims that it can provide a new metapsychology for psychoanalysis. It then becomes germane to confront Freud's quantitative and qualitative models with the explicit mathematical elaboration, in current neoclassical economics, of Bentham's rudimentary quantitative intuitions.

A recurrent concern in Freud – especially the early Freud – was building bridges between psychoanalysis and the science of his day. Psychoanalysis today tends, however, to see clinical material, presented in narrative form, as both its context of discovery and its context of justification. If we see psychoanalysis as a research programme, it is hard to resist the conclusion that its progress has become disappointing, especially given the large number of talented people in the field. Ainslie's approach allows a rethinking of psychoanalysis in which its mutual exchange with empirical, quantitative behavioural science will resume, in the process hopefully (again?) turning it into a progressive research programme. This rethinking of the foundations of psychoanalysis, if productive, will obviously not leave its superstructure unchanged.⁵

5 Wampold's (2001; 2007) meta-analyses of existing empirical research into the effectiveness of various forms of psychotherapy suggest that we need not expect the dialogue of psychoanalysis with cognitive science advocated here to lead to improved results for psychoanalysis as a therapy – according to

Ross's general reconceptualisation of the relation between economics and behavioural psychology offers a way of contextualising Ainslie's approach so that its nature, power and implications become clearer.

Reading this book as a relative outsider to economics and cognitive science was often tough⁶, but I was goaded on by the sense that the book forms a historical introduction to, and synoptic overview of, a vast tract of interesting intellectual lands (34), as well as a copiously argued, coherent position in the controversies surrounding the theoretical and philosophical mapping of these lands. Ross's book has important things to say about most of the major problems in the philosophy of the social sciences.

Ross approaches his project from the perspective of the philosophy of science. Doing this well, he says, involves 'examin[ing] the wider landscape of separated disciplinary silos in search of potentially unifying themes' (31) – and I would add, given Ross's practice: potentially unifying theses, models and explanatory procedures.

His aim is to triangulate between the findings of neoclassical microeconomics and those of cognitive science, that intersection between a congeries of disciplines including the theory of evolution, behavioural psychology, economics, AI, information theory, and neuroscience.

Faced by a choice between 'scientific' or 'humanistic' approaches to the pursuit of knowledge, he opts unapologetically for the former. In the ever increasing clash between folk psychology and behavioural science, folk psychology must yield, just as folk physics yields to the science of physics (19). It is only to be expected that folk-psychological concepts, arising as quick and dirty tools for dealing with parochial practical concerns faced in everyday life, will not identify the factors that are crucial in generating the non-obvious regularities that science as a *systematic* enterprise seeks to discover. '[R]eal causal and structural relationships' – e.g. evolutionary relationships – will often not be visible to folk observation (19), which tends to articulate reality in terms of unsystematic and contingent, ever changing human purposes.

Ross's approach implies realism regarding the existence of 'objective facts independent of any particular human purposes' (21). He sees the institutions of science as designed to detach its investigations – its search for objectively existing structures – from any *specific* human purposes (23), as well as from the vagaries of natural language. The use in science of measurable variables, and maths, generally, helps achieve this aim. It is crucial because this allows us to reach justified intersubjective agreement (23-24).

His refusal to see philosophy as independent of science places Ross at loggerheads with both the 'analytic' and 'Continental' traditions in philosophy (compare Spurrett 2008). Like Dennett, Ross believes that the philosopher 'must care primarily about facilitating the growth of scientific knowledge itself, rather than the promotion of this or that neat, preconceived philosophical 'ism'' (31). This is in line with his 'naturalism, according to which consensual scientific practice is the basic source of evidence in philosophy of science' (192). Philosophy must yield to the results of empirical science

Wampold's findings therapeutic efficacy depends not on the *content* of the theory underlying the therapy, but on other factors, such as the degree of faith therapist and client have in it.

6 I would advise newcomers to economics and the philosophy of economics to set aside a few days to read the book without interruption, so as to benefit fully from Ross's careful step by step development of his argument. (If one has difficulty recalling an element introduced at an earlier point in the text, the index usually allows one to locate the passages where new terms were first introduced, fairly quickly). The following texts can help supply the background that is needed to make Ross's book more accessible: Ross 2006; Ross 1999 (on the history of economic thought and the concept of utility); Thaler 1992; Dixit and Skeath 1999 – the latter two titles on Ross's (34) own recommendation.

where they contradict each other. As Dennett never tires of saying, pretended philosophical demonstrations of impossibilities are often no more than failures of imagination. The work of naturalist philosophers is typically an omelette of conceptual analysis and empirical theory – an omelette which cannot be unscrambled (36).

* * * * *

Here is an image of recent intellectual history that is extremely widespread just now. Several decades ago, positivism was thankfully routed by an aggrieved community of social scientists, historians, and humanist philosophers. Unfortunately, a gang of the positivists' natural associates who should have surrendered with them, neoclassical economists, hung on in a redoubt that they preserved for years through relentless obtuseness. But gradually, inevitably, the isolated fastness crumbled. Now all can see that neoclassicism has followed positivism into the dustbin of history; and the victorious humanist forces survey the broken ranks pondering whether, and if so on what terms, we should put up with any systematic economics at all.

This image deserves contestation (119).

Ross goes against the widespread tendency to pronounce neoclassical economics dead or dying.⁷ If a discipline is defined by 'its practitioners [constantly trying] to handle as many new phenomena as they can with as few new theoretical tools as possible' (7), it is incumbent on its practitioners not to call for revolutions before the resources of the *status quo* in the discipline have been exploited to the hilt. In *ETCS* this is what Ross tries to do for neoclassical (micro)economics. He is 'not telling economists that they should be going around their work in a fundamentally different way' (1-2). Rather, he wants to use cognitive science to update and defend neoclassicism (28). Central to his whole argument is the thesis that neoclassical economics, shorn of misleading interpretations, is sound science, and that the undeniable technical advances in neoclassicism have also been empirical advances – meaning that *more* phenomena can now be explained *better* (4).

Moreover, Ross argues that it is exactly in neoclassical economics that the unification of the behavioural sciences is progressively taking place. According to *ETCS*, over the past few decades the trajectories of cognitive science and neoclassical economics have been spontaneously converging.

What is economics about, according to Ross? His answer, modifying Robbins, is: 'the study of responses to scarcity by intentional systems' (148).⁸ When this is linked with Mirowski's 'big idea' – 'economics is ... about optimization of something or other by something or other' (6) we get what Ross calls the 'core neoclassical commitment to economics as the systematic science of maximization under scarcity' (29). Ross fleshes this out as follows: 'Economists, classically, treat agents as found bundles of preferences that relentlessly optimize as best they can, subject to the limitations imposed by their own information-processing and energy constraints, and to external

⁷ As can be expected, he likewise rejects the even stronger critique of neoclassicism by Dupré and others who interpret certain data 'as suggesting that neoclassicism is a self-reinforcing *ideology*' (351).

⁸ In Robbins's original formulation, the scope of economics was limited to *human* responses to scarcity – he defines economics as 'the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses' (Robbins 1935: p. 16, quoted in Ross 2005: p. 87). An essential part of Ross's project of unifying economics with the rest of behavioural science is removing this limitation of scope (87, 267), i.e. because the fit between economic theory and animal foraging behaviour – for instance – appears to be closer than that between this theory and the behaviour of whole individual human beings.

constraints imposed by their environments (including the optimizing efforts of other agents)' (26).

The better one grasps these deceptively simple ideas, the more powerful they turn out to be. They give the outline of a whole research programme in the behavioural sciences.⁹

Ross's argument, which builds on his previous work (especially Ross 1999), has a large historical component. He claims that many theses that have been taken to be fundamental commitments of neoclassicism, are in fact not that at all. He argues that nothing in the *apparatus* of neoclassical economics compels the interpretation[s] under which a variety of influential writers (e.g. Sen) have weighed neoclassical economics and found it wanting – despite the fact that the founding fathers of neoclassicism themselves often proclaimed or suggested such *interpretations* in their own writings. Ross argues that the actual advances in the development of economics were either independent of such theses, or explicitly critical of them. Many, if not most, critiques of neoclassicism are based on the readings of neoclassicism he opposes. In defending neoclassicism he is thus defending something very different from what its critics have taken themselves to be criticising! His main claims on this score are:

Neoclassical economics is not wedded to the supposition that economic agents are selfish individual maximisers, even if many of its champions have defended this idea,¹⁰ or been in two minds about it. There is nothing in mature neoclassicism that warrants the idea that it models people as inherently selfish. (Neoclassicism does not *recommend* selfishness, either (149)). Mother Teresa presents no more of a problem to neoclassicism than the person to whom *numero uno* is *numero uno*. Therefore empirical research (130-132) showing that individual people regularly display altruistic behaviour does not constitute evidence that neoclassical economics is flawed, let alone *fundamentally* flawed. Altruism and a sensitivity to social context, *pace* Sen, will be reflected in the agents' preference rankings – the common currency of revealed preference theory (RPT).

Moreover, neoclassicism is by no means committed to the idea that economics deals essentially with money and material goods. This, the layman's view of economics, is still found in certain economists. Ross (73-75) traces its pedigree back to Aristotle,¹¹ and speaks of 'the Aristotelian metaphysical psychology of the first economists' (267). Because Jevons and Marshall made an '*ad hoc* distinction between higher and lower wants', they concluded that 'human urges for beauty and goodness aren't susceptible to systematization within the field theory they favoured' (382). This semi-Aristotelianism 'supposes that the agents identified by the microeconomist constitute 'cores' around which real people build important excretions that 'rise above' the economic' (380). In this view there is a whole domain of non-material, non-monetary values about which economics must of necessity be silent.

Sen is probably the most prominent current exponent of this tradition. Though respecting Sen's technical contributions to neoclassical economics (126), Ross sees his semi-Aristotelianism as fundamentally wrong (81), and his reputed refutation of Samuelsonian revealed preference theory as without substance (139). According to

9 The 'scarcity' dimension in economics ties up with phenomena that have been addressed by Continental philosophers under the rubric of 'finitude', by Freud *inter alia* under the rubric 'Ananke' and 'reality principle', and by Buddhism under the rubric of 'dukkha' – the inevitable lack of satisfaction experienced in life.

10 Parkin's (2004, 2007) widely used introductory textbooks in economics and microeconomics, for instance, treat the self-interestedness of economic agents as axiomatic.

11 The 'humanism' Ross opposes is usually Aristotelian or semi-Aristotelian in nature.

Ross any *correct* application of Revealed Preference Theory will include all the factors that Sen sees as falling *outside* its scope. ‘Utility’, as a term of art in neoclassical economics, leaves completely open what sorts of things are valued by people; the economist’s ‘utility’ constitutes a *subjective* conception of value, which is not limited to money or material goods (87-88). Where people value other things more than they do material goods, this does not mean that we are leaving the domain of economics. Good examples of what Ross calls ‘mature neoclassicism’ are Becker (1976; 1981) – whose anthropocentrism Ross however rejects – and Ainslie (1992, 2001), whose ‘picoeconomics’ is not centrally about money and material goods.¹²

Ross thus endorses Samuelsonian Revealed Preference Theory (RPT)¹³ – to him this constitutes the core of state of the art neoclassical economics. In this he again swims against the prevailing current: ‘most philosophers of economics and commentators on economic metatheory [...] regard the Samuelsonian conception as obsolete’ (123). Ross takes his time to separately hack away at the two pillars of the standard argument against RPT: 1) it is tautologous, and a tautologous system cannot be part of empirical science; 2) people are in fact not the sort of economic agents RPT describes. Part of the reason he opts for Samuelson’s approach is that it focuses ‘strictly on the abstract mathematics of maximization among competing ends and the competing means to them, and everything that might be thought special about people disappears from view in this formulation’ (268).¹⁴ The upshot is that the expression ‘economic agent’ can apply to *any* agent – neurons, neural modules, bugs, people, firms, countries – in *any* market. It can thus play a major role in the unification of science (247ff). RPT’s ‘notion of agency ... derives the concept from influences over market processes ... rather than from typical aspects of personhood’ (247).

The preferences that are central to this definition of economic agents are not introspectible entities (as Robbins, one of the founding fathers of neoclassicism explicitly thought), but revealed in behaviour.¹⁵ Similarly there is no claim that the *maximization* studied by economics is an introspectible process.¹⁶

12 Ainslie is particularly illuminating on how expressing ‘rewards’ in money terms changes the way in which imminent and delayed rewards are weighed against each other.

13 Or more precisely, his own rational reconstruction of the Arrow-Debreu update of Samuelson’s axiomatization of RPT.

14 In Samuelson himself, however, there are still traces of the Aristotelian prejudice that the economic agents described by Revealed Preference Theory can *sometimes* be equated with *people* (109). The sparse, formal nature of the model found in RPT should become evident from the following. In RPT, after its elegant axiomatization by Debreu,

to call an agent a ‘utility maximizer’ is officially not to ascribe any internal motivational properties, but is merely an abbreviated way of saying that some ... range of its behavior is describable and predictable by means of a function over binary relations respecting asymmetry and negative transitivity (112).

15 Ross argues, persuasively, that Robbins’s 1) commitment to ordinality and the principle of diminishing marginal substitutability, as well as 2) his insistence that a science should be as general in its scope as is possible while having its theories supported by empirical evidence, clash with his (i.e.: Robbins’s) simultaneous 3) commitment to introspectionism and rejection of behaviourism. Because 3) is less integral to his conception of economics than 1) and 2), it will have to go. Such a move from a positivism founded on introspectionism to one taking a behaviourist line on empirical verification, parallels the historical trajectory of positivism.

16 Ross disentangles ‘preference’ and ‘choice’ from their folk-psychological counterparts, as follows:

An economically consistent agent is said to prefer *a* over *b* if, given a fixed income and set of marginal prices, she consumes a marginal unit of *a* rather than a marginal unit of *b*. No

Finally, the *rationality* of economic agents is not something revealed by introspection, nor does it require ratiocination. Instead, such rationality is essentially a question of consistency of preferences, as revealed in behaviour. Such rationality/consistency is not an empirical feature that can or cannot be discovered in economic agents – it is rather the criterion for delineating such agents in the first place. '[A]s a matter of logic, an *economic agent* must have stable preferences; otherwise RPT does not apply to it' (157).¹⁷ Ross summarizes the foregoing as follows: 'For the sake of maintaining a consistent foothold for systematic applications of the mathematical theories of revealed-preference analysis and game theory, economic agents are to be identified with well-ordered sets of preferences manifest in behavioural patterns' (317).

Most of neoclassicism's main critics – e.g. Sen – and defenders – e.g. Becker – share the assumption that people are economic agents, and *prototypical* ones at that. But Ross's 'central thesis' (132)¹⁸ is that whole people are not prototypical economic agents, and in this thesis he sees the key to resolving the problems in the philosophy of neoclassicism, while allowing for continuity of practice among its practitioners (215). There is nothing in the apparatus of neoclassicism – i.e. in Ross's 'austere logical reconstruction of [Samuelson's] official position' (213) – that commits us to the idea that the economic agents it describes are *persons*. In this Ross is at odds with almost all the other positions on the philosophical foundations of economics discussed in his book.

Seeing people as prototypical economic agents is a relic from folk conceptions of economics. It is not the result, Ross convincingly argues, of formulating a rigorous notion of (economic) agency and then showing that persons, in empirical fact, fit the bill. Whole people aren't single economic agents across their biography. RPT thus cannot be justified by showing that it gives a correct, direct model of how individual people behave (181). Ross rejects the conclusion that this shows that RPT is wrong *in essence*. His own, alternative conclusion is rather that the identification of whole people with Samuelsonian agents is untenable.

Ross agrees with those critics who claim that Revealed Preference Theory is essentially a system of tautologies without empirical content. However, RPT does not function on its own; the desired empirically testable theory is obtained by combining RPT with a 'maximization theory' (such as EUT – expected utility theory), which tells us what is being maximized (269).

Ross argues that empirical research shows that Expected Utility Theory cannot be taken as true, in general, for whole people. That RPT is tautologous, and therefore not empirically falsifiable, is not a *defect*.

If whole people should not be taken to be the economic agents Revealed Preference Theory deals with, where are such economic agents to be found? Ross says this is an empirical matter: a particular delineation of economic agents is justified inasmuch as it allows reliable predictions (317). This shows that insects and other asocial animals are

Robbinsian reliance on deliberation or calculation is called for in understanding what 'choice' means (104-105).

17 Perhaps this could have been formulated more clearly as follows: 'Once we have opted for Revealed Preference Theory, as a matter of logic an economic agent must have stable preferences'.

18 Elsewhere (suggesting a false dilemma?) Ross calls something else his 'central thesis':

Neoclassical economics of the Samuelsonian ... variety becomes increasingly empirically applicable, not because economists promote it but because it actually captures the dynamics that have been favoring the evolution of selves since our ancestors started talking (351).

far more exemplary of economic agency than persons are (251) – we ‘have a one-to-one mapping between the biological individual insect and a well-behaved economic agent’ (96). Why? Because application of Samuelsonian economics and game theory leads to accurate quantitative predictions of their behaviour (317) – for instance, mating and foraging behaviour – while such accuracy seldom obtains when we use these tools to predict the behaviour of whole people.

But can RPT be applied to people *at all*? ‘To show that RPT is useful we must find some real structures that are usefully measured – where ‘usefully’ means nonredundantly relevant to explanation and prediction – using coefficients and relations defined by its axioms’ (143). If whole people don’t fit the bill of being the ‘real structures’ in question, to demonstrate that RPT is nevertheless ‘useful’, we must show that something else *does* fit the bill. Ross argues that Glimcher’s work in neuroeconomics does this for certain neural structures, and Ainslie’s piceconomics for certain subpersonal structures (‘interests’ or ‘successive selves’). If Ross’s argument is sound – which it perhaps isn’t (see below) – RPT would thus be empirically vindicated as being applicable to people.¹⁹ In discussing these two approaches Ross also aims to bolster his claim that RPT, as well as the intentional stance à la Dennett, are central to the unification of the behavioural sciences.

People should usually be modelled as comprising multiple economic agents, simultaneously or successively. Only under very specific circumstances do they approximate to the economic agents described by neoclassicism. ‘People, like countries ... are, from the economic perspective, *macroeconomic* objects in the first place’ (381); they can be modelled as markets (319).

Information-flow problems in systems as complex as human beings preclude the existence of a central locus of control, and whole people are therefore not paradigmatic economic agents. Inasmuch as they *do* approximate economic agents, says Ross, this will be the result of external scaffolding. Coordination between people requires the use of language and other public signalling systems – and involves (more specifically) ‘the narrative construction of selves as behavior-stabilization devices’ (317).

This brings us to the work of Daniel Dennett, the ‘hero’ of this volume (14). Ross has played a major role in disseminating and developing the Dennettian programme – especially into the area of economics, which is not Dennett’s own *forte*.

We have already seen above that Ross sees the focus of his book as ‘the relationship between microeconomics and the theory of intentional behavior’ (15). The importance of this notion to economics is that economic agents are intentional systems, and the notion of an economic agent is central to economics. As we saw above, Ross believes that the key to solving various longstanding methodological and theoretical issues in economics lies in being clear about intentionality, and for this he turns to Dennett.

The notion of the ‘Intentional Stance’ is central to Dennett’s thought. When we take the intentional stance toward an object, we treat it as an intentional system: we assume that it is a rational agent, with the beliefs and desires one can expect, ‘given its place in the world and its purpose’ (Dennett 1987: p. 17). We then predict its behaviour by

¹⁹ In the neuroeconomic example (Glimcher), Revealed Preference Theory is combined with expected utility theory (EUT) to obtain a theory that is rich in explanatory and predictive power. In the case of piceconomics (Ainslie), it is combined with a tweaked version of Herrnstein’s matching law. In certain other cases the best fit with the data is obtained when Revealed Preference Theory is combined with prospect theory.

assuming that it ‘will act to further its goals in the light of its beliefs’ (Dennett 1987: p. 17).

The Intentional Stance is central to both folk psychology and to science – including the science of economics. In folk psychology the default pattern of inference is: ‘if a person *desires* outcome *x*, and *believes* that doing *y* is a good way to bring about *x*, and *fears* no other consequences of doing *y* to an extent that outweighs the desirability of *x*, and [correctly? AG] believes that she can do *y*, then she will do *y*’ (37).

How does folk psychology relate to economics? The economist’s way of talking about preferences and expectations is a way of making the intentional stance amenable to a mathematical treatment. ‘The neoclassical economist’s ‘preferences’ look like technical regimentations of ‘desires,’ and her ‘expectations’ seem to be beliefs with explicit probability values attached’ (37). The notion of ‘expected utility’ combines both these aspects.

Intentional-stance functionalism, Dennett’s position on intentional states, contrasts with both *internalism* and *eliminativism*. Against *internalists* he argues that intentional states are never simple givens – for both the first and the third person case they are *ascriptions*. *Eliminativists*, on the other hand, think that because drilling down – into the details of the brain – won’t reveal intentional states, we can dispense with them. According to eliminativists, beliefs and desires play no role in the causation of behaviour. They belong as little in a scientific account of the world as do witches; the ultimate scientific explanation of behaviour will contain no reference to the mental, though for the time being mental terms may serve as useful heuristic devices for conducting our lives, and even science. Dennett agrees with eliminativists that when we ‘drill down’, we won’t find any intentions, but for him this is the wrong place to look. As Ross (61) puts it: ‘Refine intentional attribution not be drilling down, but by going wide – into the social environment and into biological and cultural history’. This involves triangulation between numerous indicators, each of which is usually inconclusive on its own. For Dennett,

your belief just consists in the fact that your whole system of dispositions is such that your behavior is consistent with that belief, and inconsistent with its denial. ... [T]o have intentional states is to exhibit behavioral patterns that can’t be predicted or explained without recognition of the patterns indexed by the intentional states in question (63).

This is why eliminativism is wrong; intentionality – the notion of agency – *cannot* be eliminated without crippling behavioural science. While Ross’s position is certainly the stuff of humanists’ nightmares (190), we thus see that he actually stops one crucial step short of the humanist’s worst fear, eliminativism.

Ross’s critique of eliminativism regarding the mind, consciousness, agency and selves is central to his argument in *ETCS* (51). Although none of the phenomena are eliminated, they *are* reconceived; this critique thus does not call for a return to folk psychology.

For Ross some variety of intentional state functionalism is always going to be needed in economics, and behavioural science in general. Regardless of whether the components into which we analyse the person or agent themselves end up being intentional systems or not, we always have to *start* by taking the Intentional Stance to the person or agent in question. The only question is whether our intentional stance functionalism will turn out to have been only *methodological*, not ontological – this is the

case when the analysis starting with the assumption of intentionality ends up with components (and thus an ontology) to which intentionality is foreign – or whether it will turn out to be *ontological* as well – this is the case when even the most elementary components of our analysis still turn out to be intentional systems.

In the former case scientific progress in the study of intentional systems leads to the elimination of intentionality by showing it to be redundant. This happens whenever the Intentional Stance does not afford any explanations or predictions of the behaviour of a system that go beyond those enabled by the *design stance* and *physical stance*²⁰ on their own. An example is Glimcher's (2003) work in neuroeconomics, which *starting out* from the Intentional Stance, *ends* with the 'discharge of intentionality' – that is, with ultimate components that are *not* irreducibly intentional. However, even in such cases it is only by *starting* with the utility function of the whole organism – something that requires the intentional stance – that we can *end* up with a decomposition of its neural functions into elements where intentionality is discharged.

An example of 'ontological intentional stance functionalism', in which the ultimate components to which an explanation refers are themselves still intentional systems, and the corresponding explanation an intentional stance explanation, is Ainslie's (2001, 1991) piceconomics. We discuss Glimcher, and then Ainslie.

Neuroeconomics. A central question in Ross's book is: *is neoclassical microeconomics more than just a highly articulated formal technique? Does it have empirical application?* In Chapter 8 this is further narrowed down to the question: *given that selves – whole people – are not straightforward (or 'paradigmatic') economic agents, does microeconomics apply to people at all, and if so, how does it apply?*

Bugs are straightforward economic agents, and applications of neoclassical microeconomics to animal foraging and mate selection have been highly successful – in fact much more successful than applications to whole people. (See Glimcher 2003: p. 205-223 for an overview). Glimcher wishes to extend this programme – 'trivially', he says – to the neural level, so that *neuropsychology* starts involving *neuroeconomics*. Simultaneously he extends this research to social animals – monkeys, in his case – a prerequisite if we want our animal research to throw light on, or generate hypotheses about, humans.

The programme of neuroeconomics involves treating parts of brains as economic agents. Neuroscience would have no need for neuroeconomics if the Descartes-Sherrington-Pavlov programme of modelling the brain in terms of reflex arcs were viable. However, it isn't.²¹

Given that the bottom-up reflex arc approach won't work, Glimcher opts for a top-down approach, which typically consists of three steps: 1) determine (hypothesize)

20 The physical stance (Dennett 1987: p. 16) works as follows:

If you want to predict the behavior of a system, determine its physical constitution ... and the physical nature of the impingements upon it, and use your knowledge of the laws of physics to predict the outcome for any input.

In the design stance (Dennett 1987: p. 16-17):

one ignores the actual (possibly messy) details of the physical constitution of an object, and on the assumption that it has a certain design, predicts that it will behave *as it is designed to behave* under various circumstances.

21 Why is it not viable? Firstly, there are no, or very few, reflex arcs in the brain. Secondly, the efferent/afferent distinction presupposed by the reflex arc model breaks down at the level of neurones – only at the peripheries of the nervous system is it generally possible to classify neurones (or circuits or brain struc-

a function for a functional group of neurons; 2) specify the computations needed to fulfil this function; 3) search for experimental evidence that this computation is in fact being carried out by the neurons in question.

Glimcher's (2003) book is part programmatic manifesto for neuroeconomics, and part report of the first steps in the execution of this programme. This enterprise is easier to understand in the light of Glimcher's reading of the Cartesian programme for studying the brain. For Descartes behaviour can be parcelled out into two fundamentally different lots, the predictable and the unpredictable. The predictable part can be accounted for entirely in terms of physical mechanisms, to wit: reflex arcs. For the unpredictable part, however, we have to invoke the non-physical mind, because physical processes can only account for predictable behaviour.

Glimcher wants to overcome the Cartesian approach to the brain once and for all, by giving a physical, neurological explanation of unpredictable behaviour, that is, the non-parametric²² behaviour economists describe in probabilistic rather than 'determinate' terms. Glimcher's example involves a 'worker' playing an 'inspection game' against an 'inspector'. 'In an inspection game, one player faces a series of choices either to work for a reward, in which case he is sure to receive it, or to perform another, easier, action ('shirking'), in which case he will receive the reward only if the other player, the inspector, is not monitoring him' (326).

The optimal rate of working to shirking is a function of their relative costs and benefits. These depend on the inspector's *de facto* rate of inspecting or not, which in turn depends on the costs and benefits of inspection to the inspector. However, any detectable pattern in the worker's working/shirking behaviour will lead to a higher rate of shirk detection, and vice versa for the behaviour of the inspector. So, on a move by move basis, the actions of both the worker and the inspector must be as unpredictable – i.e. as random – as possible, though on a molar level a law-like distribution (say 50/50, or 66/33) between the two available options will be optimal to each party, thus dictating what economists call a 'mixed strategy'. For game theorists the 'solution' of a game – its so-called 'Nash equilibrium'²³ – occurs when a set of strategies is reached such that no player could become better off if any of the players were to change her strategy.

For one of Glimcher's crucial experiments on monkeys, 'working' and 'shirking' were expressed by fixating on two different points in the visual field, with the payoffs delivered as two, one or zero squirts of fruit juice.

Glimcher's main argument (sic) is that neuropsychological method should consist in using economic analysis to generate hypotheses that assign maximization

tures) as either efferent or afferent. Thirdly, reflex arcs, as determinate mechanisms, are unsuited to modelling the fundamental indeterminacy (probabilistic nature) of brain processes. This indeterminacy follows from the incompleteness of the animal's knowledge of its environment, as well as the fact that its strategic behaviour needs to be unpredictable on a move by move basis.

22 Non-parametric behaviour is behaviour that strategically takes into account the strategic behaviour of other economic agents.

23 In game theory,

[w]hat [is] referred to as [the] 'solution' [of a game] is the unique Nash equilibrium of the game. ... Nash equilibrium (henceforth 'NE') applies (or fails to apply, as the case may be) to whole sets of strategies, one for each player in a game. A set of strategies is a NE just in case no player could improve her payoff, given the strategies of all other players in the game, by changing her strategy (Ross 2006).

Many non-zero sum games have multiple NE.

problems to functional groups of neurons. These hypotheses may then be tested by monitoring brains ... while subject animals perform carefully isolated tasks (325).

Computational analysis told Glimcher that one of the tasks the monkey brain (the neurons in the parietal cortex controlling the movements of the monkey's eyes) would have to carry out in this situation, was to track the Nash Equilibrium for this inspection game.²⁴

One group of neurons was inferred to be tracking the relative utility of working vs. shirking – that is, the relative utility of looking at one point or the other. Their firing rate was independent of which eye movements were being generated, which indicates that it was not coding for behaviour. What is more, what this firing rate tracked was *relative* utility – it was the same for 4 vs. 2 squirts, as for 2 vs. 1 squirt. When the ratio between the rewards associated with each option (fixation point in the visual field) was varied, the firing rate of this group of neurons varied accordingly. Neoclassical economics was being used to explain both the monkey's gross behaviour (eye movements), *and* the behaviour of its neurons.

Moreover, applying economic theory allowed Glimcher to make sense of the behaviour of the monkey parietal cortex not only at a molar level, but also on a move by move basis. The monkey did not in advance 'know' how long a certain average rate of inspection (say 50%) would be maintained. Its brain was (thus) also sensitive to smaller fluctuations in the inspection rate, as these could signal a change in molar inspection rate. Both the gross and the neural behaviour of the monkey correlated well with an optimal response²⁵ to local fluctuations in the inspector's rate of inspection.

Glimcher emphasises the embryonic state of neuroeconomics as of 2003. The initial experimental results are very simple, while the research programme, if successful, will in time no doubt lead to a really complex picture.

Nevertheless, Ross thinks that even Glimcher's initial findings already demonstrate that standard neoclassical microeconomics can be applied fruitfully to the study of the brain; another reason to claim that it does, in fact, have empirical purchase. Because monkeys and humans don't differ in how they play inspection games, the monkey results should apply to human brains as well. This would then also show one way in which neoclassical microeconomics is applicable to humans, even if whole people are not paradigmatic economic agents.

It is now clear why Ross sees Glimcher's neuroeconomics as an example of '*methodological* intentional stance functionalism'. Although we end up with components that are less than economic agents (e.g. neurons tracking relative utility), the whole process leading to this result had to start off by initially assuming the intentional stance. The first step of Glimcher's functional analysis was to identify the utility function of a monkey – the utility of two squirts as opposed to one squirt of fruit juice for the monkey. The next step was to identify a neural circuit in the monkey's brain as an

24 The pay-offs and inspection rates had to be varied in the course of the experiment, to ascertain that the correct functions were being ascribed to particular neurons or neural circuits. The element of surprise, when what neurons have come to expect is violated in the game, is crucial to the general design of Glimcher's experiments.

25 As determined by a computer programme built around essentially the same algorithms as the inspection programme.

The neurons seemed to be reflecting, on a play-by-play basis, a computation close to the one performed by our computer ... [A]t a [relatively] ... microscopic scale, we were able to use game theory to begin to describe the decision-by-decision computations that the neurons in area LIP were performing. (Glimcher 2003: p. 317, as quoted in Ross *ETCS*, 328-329).

economic agent, with its own utility function. Only later, when the functional group of neurons in question is decomposed further, do we reach components that are no longer economic agents – e.g. a group of neurons that tracks relative utility.

Because monkeys do not have the social and cultural scaffolding required for selves, monkey findings will be insufficient to account for selves. In fact, Ross uses Glimcher as one half of a two-pronged strategy for answering the question: how does (neoclassical micro-) economics apply to people's internal dynamics? The other prong of his strategy refers to Ainslie's work in picoeconomics, which is an example of '*ontological intentional stance functionalism*'. Here the intentional stance is not jettisoned after the initial analysis, but also applies to the elements into which this analysis decomposes whole people.

This brings us to Ross's account of selves, which draws on picoeconomics – Ainslie's model of the individual as a succession of economic agents that bargain with each other, or a marketplace of subpersonal interests – but also on Dennett, whose preoccupation with the concept of the self has been a constant in his otherwise diverse work in many distinct terrains (12-13).

Whole people aren't paradigmatic economic agents. Their preference rankings change for no better reason than the passage of time (Ainslie 1992; 2001). They also do not have the centralized servosystematic architecture of simple, paradigmatic agents like bugs and other asocial animals. The loss of straightforward agency is the price paid for the distributed nature of human information processing. Distributed processing is the only way in which human brains can handle the computational complexity of the problems they face. Where does this complexity come from?

According to Ross, being a social animal leads to an exponential increase in the complexity of the (economic) problems the brain has to solve, because social animals play social games. They face *both* parametric problems – those that don't change in response to your attempt to solve them – *and* non-parametric ones – the strategic problems that arise because each player's strategy has to take account of the strategies of other players, and therefore has to change in response to changes in the other players' strategies. Solving nonparametric problems requires exponentially more computational power than solving parametric ones. In interactions between humans one and the same action can simultaneously be a move in a variety of different games against a variety of different players; this makes keeping track of what is going on and pursuing optimality (even of the sufficing kind) all the more computation-intensive.

A *sine qua non* of game playing is *game determination*. Predictable selves make the problem of game-determination tractable. In game determination the question is: what game can we play that will allow each of us to promote her own goals? (If the advantages are one-sided, the other party will walk off or throw a spanner in the works). It thus requires that each have an idea of the other's goals and preferences, the payoffs that the different possible outcomes of different games will have for the other, and thus how each player ranks these outcomes. (Remember that for Ross ranked preferences – a utility function – define the very identity of an economic agent).

Ross treats the self as a device for the 'maintenance of cohesion' (317), needed because human brains and behaviour, on their own, do not have the cohesion found in the brains and behaviour of asocial animals. Selves are a way for individual human beings to *approximate to economic agents* while still reaping the network efficiencies associated with sociality.

The self is often conceived of as a central executive agent, a homunculus sitting at the controls of the brain. Dennett will have none of it. He believes that the self exists,

but not as a central executive. For him the self is virtual; it cannot be found in the brain. Ross thus denies that persons are prototypical economic agents, without denying that they have *selves*. For him the self is a product that arises when a human brain is suitably scaffolded from outside by society, language and culture. In Dennett's Multiple Drafts Model the brain – which we know to be a massively parallel processor, not a serial one – is treated as a kind of decentralised information market.²⁶ A central executive would cause bottleneck problems, nullifying the advantages of a distributed system. The relatively simple, relatively centralised servosystematic solution found in paradigmatic economic agents is thus not available to humans.

Nevertheless stability in nature, society or the economy is possible in the absence of central control, as demonstrated for instance by ecology and the relation between prices and rates of supply.

How does the external scaffolding provided by society, language, and culture turn biological individuals into selves? Their combined workings make humans *more predictable* to each other (and themselves), so that coordination and cooperation become easier and more profitable. Society puts pressure on individuals to become more *consistent* – the essence of selfhood – both because game determination needs consistency in the players to become tractable, and because, as people become less predictable, the cost/benefit ratio of game playing deteriorates rapidly. Language supplies a public signalling system that reduces the bewildering array (or continuum) of analog options to a finite number of digital options, and defines boundaries our self-regulation must respect, *inter alia* because our behaviour is also a signalling system to other players.

Language crucially also permits *self-narration* (always with the help of multiple co-authors). '[S]elves, like fictional characters, are narrated systems of behavioral dispositions that ... imply networks of expectations in multiple types of situations' (286).

Narrating the self is not a *post hoc* description of a pre-existing entity; it is *constitutive* of the self. This self-narration again makes the person more consistent and predictable, because that is a central feature of narrative. The constraints on self-narration are much the same as those on narrating a character in fiction – considerations of consistency and plausibility are central to the whole process. (Ross says that we don't want to be *too* predictable, because that would make us boring; nevertheless predictability must outweigh unpredictability. However, once predictability or consistency ceases being the be-all and end-all of selfhood, the general tenor of his account is put under strain). Like characters in novels, selves can and do change.

We become selves by taking the intentional stance to ourselves. Just as we make sense of others by ascribing – on the basis of context and behaviour – desires, beliefs and other mental states to them, so we make sense of ourselves by ascribing such states to ourselves.²⁷ What are my beliefs? My desires? They cannot constantly change, and if they change, the changes must be motivated and intelligible. I assume that I act according to my (by and large stable) beliefs and desires, and know that oth-

26 Some economists, elaborating on sketchy suggestions first made by Hayek, treat markets as distributed computational devices: the price resulting from the distributed interactions between economic agents gives information unavailable to and incomputable by the individual agents on their own (Stix 2008)).

Hayek ... argued ... that prices aggregate information held by a group – 'dispersed bits of incomplete and frequently contradictory knowledge which all the separate individuals possess.' That knowledge is combined into a price that expresses the relative desirability of a commodity ... at a given moment (Stix 2008: pp. 30, quoting Hayek 1945).

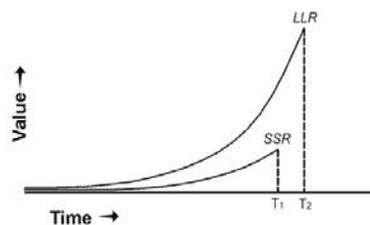
27 Dennett argues vigorously against the idea that I ascertain my own beliefs and desires by 'introspection', rather than having to ascribe them to myself much like I ascribe them to others.

ers assume this as well. The iteration of this process makes me more determinate and consistent on all these scores.

Picoeconomics. Ross's account of selves draws heavily on Ainslie's (1992, 2001) 'picoeconomic' account of persons. A good foil for Ainslie's denial that persons are straightforward economic agents is Becker's account, which goes into gymnastic contortions to preserve its assumption of 'mature anthropocentric neoclassicism': the (to Ross and Ainslie untenable) assumption that a person is a single economic agent, with an unchanged utility function, throughout her lifetime (155). Ainslie is a distinguished current representative of an opposing tradition, dating back as far as Hume and perhaps Plato, which splits the person into multiple agents.

I am saving an extended account and discussion of Ainslie's thought for another occasion – here I will be brief. Economists generally assume that people discount the future exponentially – like banks do when computing compound interest (Ainslie 2001: p. 28; 2005: pp. 636). A robust body of experimental work however indicates that this is not the case and that seeing people as hyperbolic discounters is closer to the truth. The essential difference between the two types of discounting is that, whereas in an exponential discounter a preference for A above B is time insensitive, this is not the case for a hyperbolic discounter, whose preference ranking for two items can change for no better reason than the passage of time. Crudely put in non-technical terms: for hyperbolic discounters the values attached to proximate rewards and to rewards lying further into the future do not increase proportionately with the passage of time – as the earlier reward (say, drinking on Sunday night) becomes imminent, the value attached to it increases more rapidly than the value of the more delayed reward (not having a hangover on Monday). This means that the curve plotting the value attached to a sooner smaller reward B and that attached to a larger later reward A can cross, so that while B was valued beneath A at a certain distance in time, it comes to be valued above A as it becomes imminent. By contrast, for an exponential discounter these curves will never cross: with the passage of time, the values attached to sooner and later rewards remain proportional to each other as they increase.

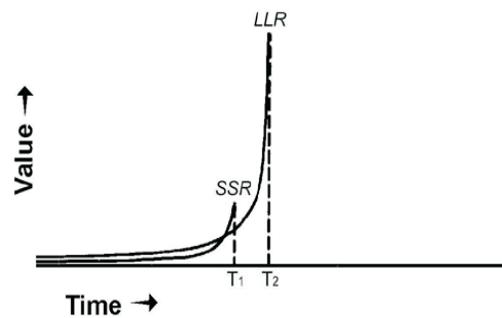
Figure 1



EXPONENTIAL DISCOUNTING

'Conventional (exponential) discount curves from a smaller-sooner (SS) and a larger-later (LL) reward', where LL reward is preferred. (based on Ainslie 2005: pp. 636).²⁸

²⁸ Figure 1A describes a situation in which the LLR is preferred over the SSR. Note that it is also possible for an exponential discounter to prefer (again: consistently) the SSR.

Figure 2**HYPERBOLIC DISCOUNTING**

‘Hyperbolic discount curves from an SS and an LL reward. The smaller reward is temporarily preferred for a period just before it’s available, as shown by the portion of its curve that projects above that from the later, larger reward.’ (Ainslie 2005: pp. 636)

We saw above that according to Ross economic agents are defined by their utility functions – their preference ranking of available commodities. An economic agent becomes a different economic agent if its utility function changes, and Ainslie shows us that many of the changes in people’s utility functions occur for no better reason than the passage of time.

A hyperbolic discounter can prefer A to B, B to C, and then C to A – this is called ‘having cycling preferences’, and has multiple disadvantages. In the first place goals aren’t achieved efficiently (or may not be achieved at all). In the second place, whoever becomes aware of cycling preferences in another person can use this knowledge to progressively divest the person with cycling preferences of all her assets – i.e., to ‘money-pump’ her – so that she can no longer influence the market. Her *agency* in the economic sphere would vanish. This serves as an argument why economic agents must per definition have stable preferences.²⁹

Ainslie reinterprets the venerable notion of ‘the will’ as the most versatile strategy for stabilizing my preferences, so that I become less subject to self-defeating behaviour. ‘The will’ makes me behave more like an exponential discounter, despite the fact that my fundamental (default) discount curve remains hyperbolic. ‘The will’ is not the name for some faculty transcending the basic mechanisms otherwise governing my behaviour – it names a ‘bargaining situation’, in which the different interests which wreak havoc by gaining the upper hand at different moments, enter into compromises with each other. The stability needed for the person to flourish is bought at the price of these compromises: none of the interests gets the payoff it would have, had it not been

²⁹ The initial results of computer simulations by Spurrett et al (2005) indicate, however, that there are some environments in which hyperbolic discounters have an advantage over similar agents with an exponential discounting function.

under pressure to make a deal with competing interests, but on the other hand most of them get a better payoff than they would have, if a rival interest had intermittently been king of the roost, and not catered for *its* rivals at all.

Ainslie's account of 'the will' dovetails with Ross's account of the self – both make human individuals more consistent than they would have been, had they not had wills or selves. In Ainslie's account the person over time is clearly not a consistent economic agent. In Ross's reading of Ainslie Samuelsonian microeconomics, which requires consistent economic agents, gets its purchase by treating each successive state of the person, characterised by a separate utility function, as a separate economic agent, and subsequently reading these economic agents as playing games with each other – games which stabilise persons over time because the collective payoff of stable compromises is higher than that of alternations between periods of absolute sovereignty and absolute powerlessness for each interest.

Conclusion

General remarks on the nature of Ross's project and argument.

In evaluating *ETCS* I start with a number of remarks about the nature of Ross's project and argument in the book.

1. *Aesthetic pleasure.* Reading Ross's book gave me aesthetic pleasure – linked in no small measure to the coherence and scope of the book, as well as the ease and economy with which the whole is orchestrated and presented – the beauty powerful theories often have.³⁰ All this is probably also tied to the fact that the book offers a map which promises to make navigating the terrain in question a lot simpler. If and to the extent that it is wrong (for instance by being *too* simple), drawing an alternative map will probably be considerably easier after *ETCS* than it would have been before.
2. *The normative and the non-normative.* On another note, it was refreshing to read a text where I had no sense that the author's normative commitments were secretly steering the non-normative parts of the discourse from behind the scenes.
3. Ross's book is a model of *academic etiquette* in its charity towards those who are criticised (no straw men here, as far as I could see) and lack of unnecessary polemical heat.
4. *Back to the foundations.* It is only when we take account of Ross's specific, and often highly original, stance re larger or foundational issues that his more specific claims become plausible. Moreover, as often happens, the writer's arguments for his claims help the reader grasp what is in fact being claimed.
5. *An anti-climax?* Chapter 9, in which Ross (among other things) again positions himself vis à vis Mirowski's five futures, is a bit of an anti-climax, given the ex-

³⁰ People's *revulsion* to approaches like Ross's often also has an aesthetic component, and perhaps legitimately so. The basic terminology of economics as the layperson encounters it, tends to be that of money, material goods, stocks, etc. In this way taking one's cue from economics causes a skewing in how nonmonetary values are thematised. Here is Ross on the midlife crisis: 'The familiar phenomenon of the mid-life crisis ... arises when people regret the formerly open possibilities their self-narratives have closed off, and so try to withdraw some but not all of their investment in their self; but, equally famously, the various pieces of their portfolio are difficult to unbundle, so valued stock is unintentionally thrown away with what's deliberately discarded ...' (294). Locutions like these give credence to the view that from the perspective of economics, paradigmatic values are indeed of the monetary or material kind that lay people take to be the full and proper domain of economics.

tent to which it repeats material already presented at the beginning of the book. But admittedly it is not the conclusion of a monograph, but only an interim summary of the state of play halfway through a two volume work.³¹ And what was initially offered as bald (or at least: hirsute) claim, has now been elaborated, and amply undergirded with argument.

6. *Fodder for anti-humanists.* *ETCS* challenges the deeply held convictions of some intellectual constituency or other at every turn. In the ambitious scope of its theses, its iconoclasm, its anti-humanism – and its difficulty – the book offers something of the sort people otherwise look for in the heroes of French theory of the late XXth Century, such as Althusser, Foucault, Deleuze, Derrida or Nietzsche.³² However, I doubt that Ross will make many converts from this constituency. On the other hand, if over the years I have myself undergone something like a conversion of this sort, others may as well.³³
7. *Opportunity for humanists.* Humanists, if they are *confident* humanists, should by rights welcome this book. Ross's book gives humanists an opportunity to sharpen and update their critique of anti-humanism. However, to many 'humanists' (like Dupré – see *ETCS* 19) Ross's book will no doubt seem to emerge from the Dark Side – embracing a position that can be shown to be wrong not only on epistemological, but also on normative grounds. Few of those in the humanities concerned with making the world a better place – with opposing injustice, overcoming poverty, racism, sexism and so on – will tend to look for the theory needed to achieve such noble aims in the places Ross does, or in *ETCS* itself. (By contrast Sen, or some variant of Marxism or postmodernism will be considered legitimate places for intellectuals of goodwill to seek enlightenment). In many constituencies in the humanities and social sciences, even *seeking* illumination where Ross does will be considered intellectually and morally suspect.³⁴ His iconoclasm occupies a very different position in the world of ideas than the celebrated, but by now domesti-

31 Ross (320) says that

[c]ompleting the theoretical construction of the economic person ... can't precede investigation of macroeconomics. Microeconomics and macroeconomics are going to be built *into* each other, not worked up as separate modules and then glued together.

If this is so, many of his substantive conclusions re the micro will have to wait until the macro has been tackled in the 2nd volume of *ETCS*.

- 32 It will come as no surprise to anybody who has read my account up to this point that the differences between Ross's approach and that of most French theory are far more striking than the parallels: where Ross uses technical language, this isn't his idiosyncratic language for his own idiosyncratic purposes, but the language of well-established fields of enquiry with technical apparatuses having a proven track record of application and formal elaboration. In Ross's case, the hard work that I needed to do in order to master the difficulty of the text had the advantage of making some major fields of enquiry, such as economics, behavioural economics and game theory, more accessible to me.
- 33 In my career as a student of philosophy I started out very much Ross's stereotypical humanist. This was later modified by a large dose of anti-humanism (à la psychoanalysis, Nietzsche, Derrida and a few pinches of Foucault), and some pomo-leaning hermeneutics (Rorty, Bernstein).
- 34 While writing this article I often discussed it informally with fellow academics, who mostly had very little knowledge of economics as a discipline. I was surprised by the confidence with which many of them believed that any defence of neoclassical economics was necessarily a defence of capitalism and the current global economic status quo.

cated³⁵ or (at least for certain constituencies) canonised iconoclasms of Marx, Nietzsche, Freud, Foucault and Derrida.

8. *Two cheers for philosophical naturalism.* *ETCS* offers a spirited defence of philosophical naturalism. For me, the simplest argument in favour of naturalism is the following: Science is our most systematic and effective truth-seeking activity. If philosophy is to be a truth-seeking activity, its default stance must thus be to take scientific method and the findings of the particular sciences very seriously indeed.

Over the last few years, I have come to reject many of the arguments or prejudices which I had against philosophical naturalism, and Ross's arguments in its defence have added to its attraction for me. Nevertheless, if I am to become a cheerleader for naturalism, then only for a fairly open-ended form of it. The naturalism Ross and Dennett promote, may itself be a broad church; think of the enthusiasm with which Rorty tended to appropriate Dennett for his own methodologically libertarian purposes. In Dennett's naturalism there is a lot of space for appropriating non-naturalists like Wittgenstein and Ryle, and even writers of fiction such as Borges and Lem (who are repeatedly anthologised in Hofstadter and Dennett 1981). It is true that Ross's naturalist church makes a somewhat narrower impression than Dennett's, but Ross nevertheless does not define naturalism in a way which would make of it a rigid orthodoxy. Moreover, in his own remarks on the self Ross is in effect implicitly addressing interlocutors of a Hegelian (Taylor 1989), psychoanalytic or postmodern persuasion on their own terrain.

While writing this review, I gradually realised that I myself won't usefully be able to walk the route Ross outlines in it. For that I should have had greater skill and facility in maths and mathematically formulated disciplines like economics, and been better at assimilating and evaluating work in fields like neuroscience and experimental psychology. In this I am probably not atypical as a philosopher. Hopefully I can watch from the periphery and find ways of appropriating highly technical work such as that engaged in by Ross, in useful ways. Also, I would want my naturalism to be broad enough to allow me to write the occasional article on art or psychoanalysis which need not prove that the writer is a card-carrying naturalist.

Another reason why I would like my naturalism to be of the relaxed variety, is that I value the fact that philosophy is one of the few academic spaces where thinking which does not fit into a particular box (or one of a small number of approved boxes), is encouraged, or at the very least, countenanced. Ordaining that all philosophers should be naturalists – where this is linked with a narrow, exclusive, prescriptive definition of acceptable topics, interlocutors, styles of argument or philosophical methodologies – would undermine this. My fear is that this would be like destroying seed banks for the preservation of genetic diversity – all those countless species of wheat and other grasses that from our current perspective seem to be of little or no value. Philosophy should not put all its eggs in one basket. It is unwise to attempt to define a single correct approach for all philosophy. Philosophy needs controversy, and the very nature of the discipline has historically been one of the most controversial topics in philosophy. Naturalism will

35 Atheism was central to the projects of Marx, Nietzsche and Freud. The way in which these writers have become *salonfähig*, even in the most religious of salons, is a good index of the degree to which they have become domesticated.

contribute more to philosophy if there are philosophers who oppose it, or investigate other, non-naturalist avenues.

Next, naturalism seems more obviously relevant to some areas of philosophy (philosophy of science, philosophy of mind, epistemology ...), and less obviously so to others (aesthetics, ethics, history of philosophy, philosophy of law). Though I am very interested, for instance, in evolutionary approaches to ethics and aesthetics, it is by no means clear that evolutionary and other naturalist approaches will lead to more interesting results here than non-naturalist ones. Perhaps these branches of philosophy are about inventing new ideas and perspectives as much as, or more than, they are about truth seeking. Would it be a fatal objection to the neo-Platonist aesthetics inspiring Piero della Francesca's paintings if it were not *true*? And as for ethics: it is not clear that philosophical naturalism requires one to be a moral realist. My rejection of slaveholding as a social practice would not become any less firm if I were convinced that it finds no support in arguments from naturalist philosophy, or even that there is no independent criterion for ascertaining that an ethics rejecting slaveholding is more *true* than one accepting it. Perhaps an ethics rejecting slaveholding should best be seen as a new invention, a cultural innovation, and perhaps advances in ethics are usually of this sort.

Reservations regarding some specific themes in ETCS

Having made these general remarks on Ross's project, I now want to express some reservations about four specific themes in *ETCS*: the nature of the self; the lifespan of successive piceconomic agents; Ross's distinction between straightforward and nonstraightforward economic agents; and, finally, the possibility of more 'revolutionary' alternatives to neoclassical economics. Due to space constraints my second and third points will be brief.

1. The self.

Ross's theory of the self is central to *ETCS*. If pressed, I would probably go for something similar. Nevertheless, I have some doubts. I start with my methodological concerns.

Ross's account of the self seems far less systematic than most of the rest of the book. (Could this lack of systematicity be a defect of presentation rather than of substance? Ross's expository skills plead against this). As Ross moves from an account of economic agency to one of selfhood there is a striking change of epistemological gear, which he does not signpost. Is this because he is insufficiently aware of it himself?

The bar he sets for his theory of self in practice seems much lower than the one he previously set in theory for anything aspiring to the quality label 'naturalist philosophy'. To me it is quite acceptable if a philosopher honing in on a new theme gropes about in the dark to some extent, but this does not seem to be how Ross conceives of it himself. In accord with the ambitions of a fairly rigorously conceived philosophical naturalism, he is not at all modest about the epistemic status of his 'theory of the self', as locutions like 'the economic modelling of human selves' (30, compare also 118), and '[a] scientific account of the self' (280) attest.

What can one demand from a theory of selfhood? What is an appropriate strategy for talking about the self? I don't know, and don't find a convincing, reasoned answer in Ross. In fact Dennett (1991a: pp. 410-411; a major source for Ross's theory of the

self) admits that the postmodernist – and thus presumably epistemologically and methodologically libertarian – theory of selfhood sent up in a quote from a David Lodge novel can be read as a caricature of the theory he himself subscribes to.

What are the phenomena a theory of the self will have to explain? What are the strengths and weaknesses of competing contemporary accounts of the self? (Ross only discusses the historically influential, but by now dead in the water Lockean-introspectionist account (225)). It is hard to see how one can develop a scientific, or even just a systematic, theory of the self without answering questions like these.

The relation of Ross's account of agency to empirical work is clear, while that of his account of the self to existing or future empirical research, isn't. (Compare for instance the many empirical claims on pp. 288-289 for which no empirical evidence is given.) While his account of agency is highly systematic, his account of selfhood tends to the 'rhapsodic' in Kant's sense – it is not clear how or even whether the various bits of it fit together.

Ross's account of economic agency does indeed seem detached from parochial concerns and folk notions, as is fitting for a naturalist philosopher. But in his account of the self the concerns guiding his theory seem to be continuous with (or not clearly and explicitly detached from and contrasted with) the parochial everyday concerns on which folk notions of selfhood are based. I must admit that I would personally be quite happy for a theory of the self which allowed me to reflect more systematically on some everyday concerns, even if something like a 'science of the self' (or an explicit economic model of the self) was still nowhere to be seen. But Ross has led us to expect something like the latter from him.

Ross's account of economic agency is securely anchored in two well-regulated discourses, both of which are clearly distinguishable from folk notions of agency: 1) neo-classical economics, with its mathematically expressed, formalised apparatus, and 2) the Dennettian theory of intentionality (if suitably updated in the light of Dennett 1991b). When it comes to his account of selfhood, however, it is not clear which connotations of folk conceptions of self are still relevant or permissible, partly because it does not seem to be clearly anchored in any discourse as well-regulated as his account of economic agency. (Game theory does not provide such an anchoring discourse, as Ross's use of game theory to update Dennett does not seem to undergird and unify his whole theory of the self).

Whereas it is clear that Ross's notion of economic agency is a very 'thin', highly abstract and formal one, out of which all 'thicker' folk connotations have carefully been distilled, it is not clear how thick his notion of selfhood is. Ross foregrounds the self as a device for increasing the consistency, cohesion and predictability of an organism that is not naturally a paradigmatic economic agent. But far more seems to be involved – the impression is often that the whole gap between the 'thin' concept of an economic agent and the 'thick', multi-dimensional concept of a person, is supposed to be bridged by the notion of a 'self' – at one point Ross (318) for instance makes selves 'ontologically equivalent to real people'. 'Selves' must thus apparently be taken to be real people in all their complexity. If this is so, formulating a theory of the self becomes a dauntingly complex task. Perhaps no elegant, 'scientific', 'economic' or even systematic model of 'real people' is possible.

If we want a 'thick' model of the person, we would have to bring in the body (this would perhaps already be needed for even a very simple model), as bodily continuity is still the crucial, foundational, default criterion of ('personal') identity, and my body

is central to my perspective on, my orientation in, and my navigation through the world. (Presumably all the other features I ascribe to my self are in some way grafted onto my body schema, or at the least my body as address for the self). We could also ask: what about gender? Ross's rationale for leaving out exactly the dimension which in everyday life is usually the first thing we want to know about another person is unclear to me – could it be the fact that 'gender' simply is not part of the basic concepts of economics? But then economics would by that token not be a sufficient basis for a theory of the self.

Perhaps Ross's argument in *ETCS* does not actually need a fully-fledged theory of the self (if such could be had), and he should rather restrict himself to his argument for the claim that selves make people more consistent; more like economic agents or exponential discounters than they otherwise would be. This is a 'thinner' issue, less like a theory of everything, a theory of 'real people' in all their complexity.

Ross's use of the term '*self-narration*', and the *fictional metaphor* of self-construction also needs critical scrutiny. 'Self-narration' is a, or perhaps *the*, central term in the Dennett-Ross theory of the self. In the absence of a more formal, systematic theory, we are doubly exposed to the vagaries of natural language, so that the connotations of our terminology become highly important. They determine how the term works as (part of) an intuition-pump. I think that the term 'self-narration', and the fictional metaphor Dennett and Ross associate with it, pumps the reader's intuitions in all sorts of wrong, and probably unintended, directions.

The term 'self-narration' suggests reflexivity, that is, that the entity doing the narrating is also the entity being narrated. This automatically demotes the importance of the contribution of co-narrators to the second rank. (Other locutions suggest similar connotations: 'according to Dennett *human selves create themselves* by taking the intentional stance' (160 – my italics)).

'Self-narration' also suggests that the verbal behaviour which is constitutive of the self, consists of narrations. However, *all* verbal behaviour is (potentially) constitutive of the self, i.e. signals to others and myself who and what I am, and can be expected to be or do in future. And not all 'bookkeeping' (389) of verbal and non-verbal behaviour takes the form of narrative. (Think of items on a CV (résumé), IQ and other test scores, or criminal records).

Moreover, the term 'self-narration' obscures the degree to which (potentially: all and any) non-verbal behaviour is constitutive of the self (in the sense just glossed). Capitalism for instance would not be capitalism, and marketing not marketing, if consumer choices were not taken as crucially constitutive of self or 'identity'. This is not to deny that the categories in terms of which non-verbal behaviour is interpreted, are crucially linguistic, regardless of whether this takes the form of 'narratives' or (some other form) of 'bookkeeping'.³⁶

Furthermore, the term 'self-narration' and the related fictional metaphor obscure, or at the very least do not foreground, the constraints on current self-narration imposed by past self-narration, which in the case of a self is never freely or consciously started or chosen by the individual in question. Ross (281) and Dennett correctly emphasise that Conan Doyle, telling a *new* episode in the life of Holmes, is constrained by decisions made in telling previous episodes. However, the analogy between self-narrating

³⁶ When it comes to what *I* do to send signals about my self, 'self-presentation' would perhaps be a more neutral term to use than 'self-narration', as it also covers non-narrative verbal behaviour, as well as non-verbal behaviour.

and Conan Doyle breaks down inasmuch as with Doyle this was not always the case. When Doyle started (281), highly generic narrative constraints were the only constraints, and Holmes could have become any of a whole range of possible types of detective, a Black or female detective, or not a detective at all. (A better fictional model than Conan Doyle would be an author who never had a free hand, but from the beginning had to continue an already extant series – say the James Bond series – that had been developed at length and in detail by a previous author or authors).

Finally, the fictional metaphor obscures the way self-narratives are constrained by *non-narrative* facts. It is clear that apart from the way in which in real life future narrations of the self are constrained by past ones, there are countless *non-narrative* constraints on this narrative, usually called *facts* about oneself and one's context. Think of how the stories I can plausibly tell about myself in any given social context are constrained by my gender, sexual orientation, race, class or caste of birth, temperament ('personality'), degree of physical attractiveness, athleticism, physical prowess or intelligence. (Of course one can lie about, or distort, facts like these, but because this will come at a price, and can even backfire completely, it is still true that one's narration is constrained by them). Ross himself admits that while narratives are constructions, '[m]any constructions ... are not interestingly avoidable' (280). However, this isn't at all what the fictional metaphor would immediately suggest. Perhaps 'spin-doctoring' would be a better term/metaphor than 'fiction' (*tout court*) for how one, in presenting one's self to oneself and others, 'puts a spin' on all those facts which can't be made to go away completely, including facts about how one's self has hitherto been narrated.

To summarise the foregoing: 'self-narration' as basic metaphor for the constitution of the self is full of connotations that are untenable, and to which Ross and Dennett would probably not explicitly want to subscribe.

2. *The lifespan of successive piceconomic agents.*

The plausibility and usefulness of Ross's utilisation of Ainslie's notion of game playing between 'successive selves' seems to require that each successive agent will last for a nontrivial length of time. What can we expect the lifespan of each successive agent to be? All else being equal, the more items (at different delays) a utility function ranges over, the shorter will be its lifespan – the lifespan of the economic agent constituted by this utility function. (The more items there are whose discounted values over time we plot, with one curve per item, the more often these curves will cross. The more often these curves cross, the shorter the average period will be in which a utility function – a preference ranking – will remain unchanged). A utility function ranges over the whole set of items about which an economic agent has preferences. Now this set is vast. Therefore the ranking will change all the time. Accordingly, changes of utility function will occur at time scales approaching the infinitesimal, successive agents will proliferate and there need never even be a recurrence of the same economic agent. In the light of this, it is not clear that Ross's approach will in fact salvage the practical applicability of RPT to real people, as was his intention.

3. *Ross's distinction between straightforward and nonstraightforward economic agents, or: is approximation the only game in town?*

Ross argues that we should expect neoclassical economics to apply to whole persons at best only in a very approximate way, as they are not paradigmatic economic agents.

The thrust of his argument, and many of his specific locutions, suggests that there are other entities – bugs, Glimcherian neuronal circuits, or Ainsliean ‘successive selves’ – to which neoclassical economics can correctly be applied, in a straightforward, rather than just an approximate way. However, the reservations expressed in the foregoing two points give reason to think that even in these cases the application will be at best approximate, and more reasons can be added:

- a) For one of his important claims (regarding the ability to predict the monkey’s moves on a play by play basis) Glimcher just asserts a ‘correlation’, without quantifying the *degree* of correlation (328; Glimcher 2003: p. 316-317).
- b) There does not seem to be any single version of foraging theory (to which Ross and Glimcher appeal) that accounts equally well for data from disparate sources (Rachlin 1989: p. 180; this is apparently still true twenty years later). Moreover, one of the main predictions of foraging theory, the zero-prey rule, has not been experimentally confirmed (Glimcher 2003: p. 220).
- c) Ainslie’s theory is *based on* good, solid empirical work, but has not, or hardly, been independently empirically validated, and Ainslie (2001: Ch. 8) himself admits that such validation may even in principle be elusive.

4. ‘Revolutionary’ alternatives to neoclassical economics.

It often seems as if due to the influence of caricaturised popularisations of Kuhn and Popper every Tom, Dick and Harry now hopes to instigate the next scientific revolution. (Which according to the caricature is being held back by little more than our uncritical complacency regarding the disciplinary status quo). In this context Ross’s emphasis on the importance of continuing tenaciously with normal science as long as that is possible is certainly salutary. Calls for a revolution may sound very sexy, but if normal science turns out to be able to handle the apparent anomalies in question quite adequately, the sexy would-be revolutionaries are going to have egg on their face. (In the case of neoclassical economics the called for disciplinary revolution is often – perhaps even usually – believed to go hand in hand with the social revolution that would become possible once the current economic dispensation ceases having an ideological legitimation in the so-called science of economics in its present form).³⁷

However, the strong arguments in favour of exploiting the resources of normal science to the hilt before calling for a revolution, may blind us to the need to hedge our bets between efforts to advance normal science by puzzle-solving business as usual, and efforts to explore revolutionary alternatives. Kuhn gives us reason to think that ‘hedging our bets’ is the appropriate phrase here. Philosophers of science at one stage hoped to discover some criterion of correct scientific procedure or rational scientific theory choice (such as a crucial experiment) which would allow us to determine quickly, and once and for all, which of any two rival scientific theories was to be preferred.

Rather than such instant decidability, Kuhn says that there is usually a protracted period in which good (‘rational’ and ‘scientific’) reasons can be given for both sides in such rivalries, during which it is by no means clear which way the weight of argument leans. For any individual scientist, choosing for one side or the other is then at the best of times a calculated gamble. While individual practitioners usually have to choose – if they want to contribute anything useful – science as a whole needn’t; it *can* and perhaps *should* hedge its bets. (Rough parallels exist with the argument *against* a com-

³⁷ See footnotes 7 & 34.

mand economy, and *for* an economy in which risk-taking entrepreneurs try to bring competing products to a market).

For our present purposes this means that even though I have rehearsed Ross's arguments and pronounced them to be impressive, I realise that rival approaches need not be wrong or without merit.

I touch briefly on one argument for the necessity of a more 'revolutionary' refashioning of economics than Ross's 'reconstructive' approach – the argument given by Smith and Foley (2008) in 'Classical thermodynamics and economic general equilibrium theory'. I discuss their approach as just one in a family of related approaches which rely on systems theory.³⁸

Smith and Foley take as their point of departure a formal comparison of neoclassical economics to physics. Mirowski (especially Mirowski 1989) and others have investigated the long history of attempts to model economics on physics, extensively. Samuelson (quoted in Smith and Foley 2008: pp. 9) argued that despite the many striking parallels that exist between classical thermodynamics and mathematical economics, it was misguided to expect each of the central concepts in economics to have an analogue in physics. Each was *sui generis*, so that economics would not be illuminated by pursuing these analogies further. Smith and Foley disagree. While acknowledging numerous fundamental errors in the ways this relation has been conceived of historically (starting with Walras), they think that a clear-eyed review of the analogies will enable us to identify and address some fundamental problems in neoclassical economics precisely. They accordingly set about mapping the parallels between classical thermodynamics and neoclassical economics in a systematic way, using a formalism that will make it easier to compare like with like.

This sets the stage for their claim that neoclassical economics is, like classical thermodynamics, a theory geared to systems in equilibrium, and that like classical thermodynamics, it cannot be generalised to cover non-equilibrium systems. Just as in physics we needed a new non-equilibrium theory of which the equilibrium systems described by classical thermodynamics turned out to be a special case, so in economics we need a yet to be devised non-equilibrium theory, of which the equilibrium case described by neoclassical economics would just be a particular case.

Smith and Foley claim that *in practice* neoclassical economists *do* behave as if their model can be extended unproblematically to cover non-equilibrium cases, proceeding as if even in non-equilibrium situations determinate values can still be attached to market price, welfare functions, and the other notions neoclassical economics needs if its apparatus is to have empirical purchase. For Smith and Foley the mathematical parallels between classical thermodynamics and neoclassical economics show that this is not the case: in non-equilibrium systems the values in question become indeterminate. They concede that neoclassical economics can still more or less work for *one* class of non-equilibrium economic systems, Gorman systems which, unusually for non-equi-

38 Their approach partly overlaps with a line of argument John Collier has been toying with, and discussing with a number of his colleagues at UKZN, myself included, over several years. Whereas Smith and Foley focus on the non-extendibility of neoclassical economics to (most) nonequilibrium systems, Collier's main worry is that the concept of equilibrium in neoclassical economics is less than clear; if we take it to be defined by the notion of a Nash equilibrium, multiple equilibria are possible. If this is so, the analogy with physics breaks down, because statistical mechanics proceeds on the assumption that the equilibrium for a system is unique. To put it in a nutshell: for Smith and Foley neoclassical economics runs into problems once we get out of the rather special case of equilibrium economics; Collier agrees, but for him it is already problematic what the special case of equilibrium itself would be.

librium systems, are neither path-dependent nor irreversible. However, most non-equilibrium systems will be path dependent, irreversible (and completely non-aggregatable). These features, they contend, are fatal for an extension of neoclassical economics to non-equilibrium systems.

I am not qualified to judge the merits of these claims, but Smith and Foley's arguments seem to deserve serious consideration. Though they do not address Ross's analysis explicitly, Smith and Foley implicitly oppose it. Ross says, in effect, that once we have identified suitable entities to count as Samuelsonian agents, and a suitable 'maximisation theory' (269, etc.) to combine with RPT in any particular case, the apparently insurmountable problems regarding neoclassical economics will disappear (presumably just leaving scientific puzzles, the standard fare of normal science). Smith and Foley's implicit critique would be: no amount of tweaking as to the identification of candidates to serve as Samuelsonian agents or the identification of suitable maximisation theories, will solve the fundamental problem that neoclassical economics essentially describes an equilibrium system; that as such it necessarily breaks down once we try to extend it to (most) non-equilibrium systems; and that the systems for which it breaks down are exactly the *normal* case with which economics has to deal. Their diagnosis for the limited empirical success of neoclassical economics, up till now, therefore differs fundamentally from Ross's.

* * * * *

I leave my evaluation of *ETCS* at this tentative overview of what to me seem to be some of its pros and cons. Obviously, Ross's main theses will be controversial. Nevertheless, even as an outsider I feel no compunction in concluding that *ETCS* is a brilliant piece of work. (This judgement could be justified even if all of Ross's main theses turn out to be false). Nothing indicates that any of the dominant positions criticised by Ross was based on a consideration, followed by a considered rejection, of anything even approaching the view he presents in *ETCS*.

The empirical base that Ross adduces for his view, is slender. To say so in the context of this article, is to emphasise the programmatic nature of the vision Ross develops in his book. There is always a radical mismatch between the detailed argument that would be needed to rationally convince sceptics of the merits of an ambitious programme such as this one, and the sketchiness of any argument that could possibly be given within the confines of a single monograph – especially one outlining a research programme still in its infancy. Ross's book is no exception. Accordingly, anybody buying into an argument like Ross's could always with some justification be accused of gullibility or engaging in an act of faith.

As is always the case with programmes in philosophy or science, the attempt to carry out the programme Ross outlines in *ETCS* will reveal difficulties (and hopefully: possibilities) not envisioned by its author. Whatever the outcome, I would be surprised if with time this book does not turn out to have changed the terms of the debate in the philosophy of economics and some related fields.

I look forward to the second volume of *Economic theory and cognitive science*, which will be devoted to macroexplanation. What has been said in the first volume will doubtless get a very different complexion once Ross gives us his bigger picture. I am curious whether Ross will remain as sanguine about economics in the aftermath of the crash of 2008, which has widely been interpreted as showing the bankruptcy of

economics as a discipline.

References

- Ainslie, George. 1991. *Picoeconomics: The strategic interaction of successive motivational states within the person*. Cambridge: Cambridge Univ. Pr.
- Ainslie, George. 2001. *Breakdown of will*. Cambridge: Cambridge Univ. Pr.
- Ainslie, George. 2005. 'Précis of Breakdown of will'. *Behavioral and Brain Sciences* 28, 635-673.
- Becker, G. 1976. *The economic approach to human behavior*. Chicago, Ill.: University of Chicago Press. (1976; 1981)
- Becker, G. 1981. *A treatise on the family*. Cambridge, Mass.: Harvard University Press.
- Davis, J. 2003. *The theory of the individual in economics*. London: Routledge.
- Dennett, D. 1987. *The intentional stance*. Cambridge, Mass.: MIT Press.
- Dennett, D. 1991a. *Consciousness explained*. Boston, Mass.: Little Brown.
- Dennett, D. 1991b. 'Real patterns'. *J. of Philosophy* 88 (1), 27-51.
- Hofstadter, D. and Dennett, D. 1981. *The mind's I*. New York, Basic Books.
- Dixit, A. and S. Skeath. 1999. *Games of strategy*. New York: Norton.
- Dupré, J. 1993. *The disorder of things*. Cambridge, Mass.: Harvard Univ. Pr.
- Dupré, J. 2001. *Human nature and the limits of science*. Oxford: Oxford Univ. Pr.
- Freud, S. 1975 (1895). 'Project for a scientific psychology'. *Standard Edition*, 1, 295-397. London: The Hogarth Press and the Institute for Psychoanalysis.
- Glimcher, P. 2003. *Decisions, uncertainty and the brain*. Cambridge, Mass.: MIT Press.
- Gouws, A.S. & Cilliers, F.P. 2001. 'Freud's 'Project', distributed systems, and solipsism'. *South African Journal of Philosophy*, 20 (3): 1-21.
- Hayek, F.A. von. 1945. *The use of knowledge in society*. *American Economic Review* 35 (4): 519-530.
- Ladyman, J., D. Ross, J. Collier and D. Spurrett. 2007. *Every thing must go: Metaphysics naturalised*. Oxford: Oxford Univ. Pr.
- Mirowski, P. 1989. *More heat than light: Economics as social physics; physics as nature's economics*. Cambridge: Cambridge Univ. Pr.
- Mirowski, P. 2002. *Machine dreams: Economics becomes a cyborg science*. Cambridge: Cambridge Univ. Pr.
- Parkin, M. 2004. *Microeconomics* (7th Ed.). New York: Addison Wesley.
- Parkin, M. 2007. *Economics* (8th Ed.). New York: Addison Wesley.
- Rachlin, H.D. 1989. *Judgment, decision, and choice: A cognitive/behavioral synthesis*. New York: W.H. Freeman and Co.
- Robbins, L. 1935. *An essay on the nature and significance of economic science*. 2nd Ed. London: Macmillan.
- Ross, D., A. Brook and D. Thompson (Eds.). 2000. *Dennett's philosophy: A comprehensive assessment*. Cambridge, Mass.: MIT Press.

- Ross, Don. 1999. *What people want: The concept of utility from Bentham to game theory*. Cape Town: University of Cape Town Press.
- Ross, Don. 2005. *Economic theory and cognitive science: Microexplanation*. Cambridge, Mass.: MIT Press.
- Ross, Don. 2006. 'Game Theory', *The Stanford Encyclopedia of Philosophy* (Spring 2006 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/spr2006/entries/game-theory/>>.
- Ross, Don, C. Sharp, R. Vuchinich & D. Spurrett. 2008. *Midbrain mutiny: The pico-economics and neuroeconomics of disordered gambling*. Cambridge, Mass.: MIT Press.
- Smith, E. and Foley, D.K. 2008. 'Classical thermodynamics and economic general equilibrium theory'. *Journal of Economic Dynamics and Control*, 32 (1), 7-65.
- Spurrett, D. 2005. 'Preference cyclicity and money pumping in a virtual economy' (written with John McCoy, Benjamin Murrell and Don Ross), presented at the Economics Society of South Africa conference, Durban, September 2005.
- Spurrett, D. 2008. 'Why I am not an analytic philosopher'. *South African Journal of Philosophy*, 27 (2), 153-163.
- Stix, G. 2008. 'When markets beat the polls'. *Scientific American* 298 (3), 26-33.
- Taylor, C. 1989. *Sources of the self: The making of the modern identity*. Cambridge: Cambridge Univ. Pr.
- Thaler, R. 1992. *The winner's curse*. New York: Free Press.
- Wampold, B.E. 2001. *The great psychotherapy debate: Models, methods, and findings*. Mahwah, N.J.: Erlbaum.
- Wampold, B.E. 2007. 'Psychotherapy: The humanistic (and effective) treatment'. *American Psychologist* 62 (8), 855-873.
- Watson, A.S. 1958. 'Freud the Translator: Some Contacts with Philosophy'. *International Journal of Psychoanalysis* 39, 326-327.