Journal of Economic Methodology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rjec20

Reflexivity, complexity, and the nature of social science

Eric D. Beinhocker[a]
[a] Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, Oxford, OX2 6ED, UK
Published online: 13 Jan 2014.

Journal of Economic Methodology, 2013
Vol. 20, No. 4, 330–342, http://dx.doi.org/10.1080/1350178X.2013.859403

Reflexivity, complexity, and the nature of social science

Eric D. Beinhocker*

Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, Oxford OX2 6ED, UK

(*Received 24 September 2013; accepted 17 October 2013*)

In 1987, George Soros introduced his concepts of reflexivity and fallibility and has further developed and applied these concepts over subsequent decades. This paper attempts to build on Soros's framework, provide his concepts with a more precise definition, and put them in the context of recent thinking on complex adaptive systems. The paper proposes that systems can be classified along a 'spectrum of complexity' and that under specific conditions not only social systems but also natural and artificial systems can be considered 'complex reflexive.' The epistemological challenges associated with scientifically understanding a phenomenon stem not from whether its domain is social, natural, or artificial, but where it falls along this spectrum. Reflexive systems present particular challenges; however, evolutionary model-dependent realism provides a bridge between Soros and Popper and a potential path forward for economics.

Keywords: reflexivity; complex adaptive systems; evolutionary epistemology; model-dependent realism; philosophy of social science; economics

Jel Classification: B41; B31; B50

1. Introduction

> I, however, believe that there is at least one philosophical problem in which all thinking men are interested ... *the problem of understanding the world – including ourselves, and our knowledge, as part of the world.*
>
> – Karl Popper, *The Logic of Scientific Discovery*

George Soros's mentor and inspiration Karl Popper wrote the above in the Preface to the first English language edition of his great work *The Logic of Scientific Discovery* (1959, p. xviii). Popper himself gave emphasis to the last part of the quote (which I have shortened) with italics. We cannot understand the world and ourselves separately from it – human beings and the knowledge they have – are intrinsically part of the world they are trying to understand. This embeddedness of human beings in the system they are seeking to understand, the limits to their knowledge of that system, the impact of their actions on the system's path, and the self-referential circularity this inherently creates all lie at the heart of George Soros's concept of reflexivity (Soros, 1987, 1998, 2000, 2006, 2008, 2009, 2010, 2012, 2013).

Unfortunately, however, Popper was wrong about one thing – not all 'thinking men' seem to be interested in this problem – not least of all economists. For roughly the past 140 years economics has taken another path. One that views the economy as an idealized mechanistic system and humans as detached rational analysts of that system rather than as embedded imperfect participants. This view claims to be 'scientific' but as Soros

*Email: eric.beinhocker@inet.ox.ac.uk

(2013, pp. 315–320) argues it fundamentally misunderstands the nature of science – in particular social science – and the nature of the economy itself.

But in the wake of the 2008 financial crisis, there is now a growing community of economists and other scholars with an interest in the real functioning of the economy and a willingness to deviate from orthodox theory in order to attempt to explain that reality. Understanding the inherent reflexivity of the economy and the fallibility of the human beings that populate it must be a part of any post-crisis effort to reform and modernize economics.

In this essay, I will attempt to build on Soros's ideas to further specify what is meant by reflexivity, situate the concept in recent thinking on complex adaptive systems, and explore the fundamental issues it raises for epistemology and social science.

2. What are reflexive systems?

As Soros acknowledges, he did not introduce the term 'reflexivity' – it has a history of use in philosophy, sociology, and economics. The term has generally been used to describe processes where an observer is also a participant in a system and there is a two-way feedback between the participant/observer and the system. Soros's own definition of the terms reflexivity and fallibility is provided in his contribution to this symposium (Soros, 2013, pp. 310–311).

In this section, I will build on Soros's articulation and attempt to provide a definition that is both general and specific, situates reflexivity with other concepts such as cybernetic and complex adaptive systems, and helps distinguish reflexive systems from other kinds of systems.

2.1 Necessary conditions

I would propose that in order for a system to be 'reflexive' it must have the following elements:

- *Environment*: There is some setting or environment; it could be a physical environment, a socially constructed environment, or an artificial environment (e.g. in a computer).
- *Agent*: There must be at least one agent interacting with that environment and possibly multiple agents interacting with each other (for simplicity I will refer to agents in the plural as that is the more typical situation).
- *Goal*: The agents must have some goal or goals they are pursuing in that environment.
- *Cognitive function*: The agents must have some way of receiving information about their environment, perceiving the state of that environment, comparing that perceived state against the goal state, and identifying gaps between the perceived state and the goal state; Soros calls this the 'cognitive function.'
- *Manipulative function*: The agents must have some way of interacting with their environment in order to change or manipulate that environment in pursuit of their goal; Soros calls this the 'manipulative function.'
- *Internal model*: Each agent contains an internal model that connects its cognitive and manipulative functions; that model contains a mapping between states of the environment and possible actions and consequences.

This last point requires some further elaboration. In reflexive systems, agents have an internal model of their environment that enables them to move from perception via the cognitive function to action via the manipulative function in pursuit of their goal. This enables an agent to posit statements like '*If I perceive state A* (cognitive function) *and take action X* (manipulative function) *then state B will result, bringing me closer to (or farther from) my goal G*'. As we will see shortly, the inevitable flaws and shortcomings in any such model lead to the particular dynamics of reflexive systems.

The above list provides the *necessary but not sufficient* conditions for reflexivity. But it can also describe any simple feedback and control system. For example, a room thermostat can be thought of as an agent with a goal of keeping a room at a certain temperature, it perceives the state of the room through a thermometer, and takes action turning the heat on and off. Such dynamic feedback control systems, which I would *not* characterize as reflexive, have been well studied by scholars of cybernetics (Umpleby, 2007, 2010; Wiener, 1948) and system dynamics (Forrester, 1961; Sterman, 2000).

2.2 Distinguishing characteristics: internal model updating and complexity

There are two additional elements that I would argue distinguish a reflexive system from a dynamic feedback system:

- *Internal model updating*: The internal decision model of the agents is not fixed, but itself can change in response to interactions between the agent and its environment; thus, there is a feedback between the perception of the environment and the agent's internal model.
- *Complexity*: The system in which the agent is embedded in is complex in two senses:
  the system has *interactive complexity* due to multiple interactions between heterogeneous agents and
  the system has *dynamic complexity* due to nonlinearity in feedbacks in the system.

When these two elements are introduced, the system is no longer a simple dynamic feedback system, but it becomes a reflexive system. The thermostat in a room has neither internal model updating nor interactive or dynamic complexity. There are two reasons why these factors significantly change the nature of the system. First, when the internal model of the agents is no longer fixed, but can change in response to changes in the environment, then predicting future states of the system depends not just on understanding the rules of the internal model, but on fully understanding the process by which those rules update and accurately predict rule changes over time. In order to predict the future path of the system, we must know something about the agents' goals and internal model – the internal model cannot just be a black box.

I would further interpret Soros's definition to add that not only can models update through changes in model parameters (e.g. Bayesian updating), but also the rules or structure of the model itself might change [e.g. Holland, Holyoak, Nisbett, and Thagard (1986) provide an example of an updating model incorporating structural change, and Hommes (2013) shows experimentally how people can use mean-reverting rules in some circumstances, trend following rules in others]. It is important to note that in some cases an agent might 'learn' and its internal model might improve its performance in mapping perceptions and actions toward achieving the agent's goals. But it is also possible that model updating can worsen the performance of the model as well.

Second, rather than the simple environment of a single thermostat in a room, we now assume a more complex environment. Imagine if the room has multiple thermostat/heating systems, each with its own temperature goals and internal programs. The system would have multiple, interacting, heterogeneous agents – what might be called interactive complexity. Predicting the future path of the temperature in the room would suddenly become much more complex – one would have to know a lot about each agent thermostat, its goal, its internal program, the heating system it controls, time delays between actions by the thermostat and responses in temperature, etc. All sorts of behaviors would now be possible – the system could reach a stable equilibrium temperature, or it could oscillate between differing temperature goals, or it could engage in some complex nonperiodic pattern of temperature swings, or it might even exhibit chaotic behavior.

Likewise, if one or more of the thermostats had a rule where the room temperature was a nonlinear function of past temperature, this would introduce dynamic complexity into the system and make its future behavior even more difficult to predict. For example, if one of our thermostats had the internal rule $T_{t+1} = cT_t - cT_t^2$, where T_{t+1} is the room temperature for the next time period, T_t is the current temperature setting of the thermostat and c is a constant, then the system would have a very interesting range of behaviors. This simple nonlinear dynamic equation is known as the logistic map. Depending on the initial temperature and the value of the constant c, the system could do anything from going to a steady state, to regular oscillations, to swinging around chaotically. Such a system is sensitive to both initial conditions and path dependent. The key point is it is difficult or even impossible for an outside observer to predict the future path of a system with such a nonlinear relationship without both knowing the rule, its parameters and the initial conditions.

3. Limits to knowledge and fallibility

The introduction of internal model updating and interactive and dynamic complexity then creates the central tension that defines reflexive systems: Internal model updating enables agents in the system to learn, but interactive and dynamic complexity makes it very difficult for agents in such a system to learn successfully.

It is this central tension that makes Soros's notion of fallibility inevitable. In order for an agent to successfully pursue its goals, it must have an accurate enough internal model of the system it is operating in so that its perceptions via its cognitive function lead to the correct actions through its manipulative function to move closer to its goals. But constructing such an accurate internal model and improving its performance through learning in a complex environment runs into fundamental limits to knowledge issues.

3.1 Flawed models in a complex world

If we assume that the agent or agents are not simply given the correct model of their environment a priori, but must discover, learn, or evolve their internal model through some process of observing data in the environment, then it is likely that the model the agent discovers will be flawed or inherently limited in some way. As Soros (2013, p. 310) notes, fallibility is core to his theory of reflexivity. Mathematicians and philosophers have discovered a number of results that fundamentally limit the knowledge that agents situated in complex systems can attain:

- *Difficulty discovering the correct model from finite data*: Agents only have access to finite amounts of data, but for high-dimensional systems there may never be enough data to discover the true underlying model.
- *Lack of knowledge of initial conditions and parameters*: Even if one has the correct underlying model, it may be impossible to predict the path of a nonlinear dynamic system without perfect knowledge of its initial conditions and key parameters (e.g. unknown Lyapunov exponent).

- *Inability to predict with finite computing time*: For some systems, there is no short cut or compressed model that allows forecasting future states without an infinite amount of computing power, one can only let the system run and play out over time (e.g. N-P complete problems).
- *No 'God's-eye view'*: It may be impossible for an agent embodied within the system to access information an agent outside the system with a 'God's-eye view' would have – this is related to Gödel's famous incompleteness theorem.

In any nontrivial setting of minimal complexity – i.e. almost any real-world economic situation – the true underlying model will not be given a priori to the agents, knowledge of initial conditions and parameters will be limited, data will be finite and noisy, computing power will be finite, and the agents doing the observing will also be participants in the system. This means that in almost any real-world situation the internal models of agents *must* be fallible. And that fallibility is also part of the system that the agents are trying to understand. We thus have the recursive loop that is at the center of Soros's concept: fallible agents try to understand and act in an environment of fallible agents trying to understand and act in an environment of fallible agents trying to understand .... Predicting the future path of such a system then requires perfect knowledge of the agent's own fallibility and the fallibility of all other agents – perfect knowledge of fallibility is truly a contradiction in terms.

Yet, it is precisely these limits to knowledge and the fallibility that they imply that the rational expectations hypothesis (REH) in economics assumes away (Lucas & Prescott, 1971; Muth, 1961). Under REH, agents live in a simple world without either interactive or dynamic complexity and with homogeneous and predictable agent model updating. Some may argue that REH incorporates reflexivity because it models two-way feedback between agent beliefs and the world. But REH does not incorporate any fundamental limits to knowledge in complex environments, the inevitable fallibility and heterogeneity such limits introduce, and the deep indeterminacy that reflexive interactions between fallible agents create.

3.2 Complex reflexive systems

In the above three sections, I have listed the necessary and sufficient ingredients for a reflexive system. I have briefly described how such a system differs from a cybernetic or dynamic feedback system. But one might ask whether this description differs from what are referred to as 'complex adaptive systems' (e.g. Anderson, Arrow, & Pines, 1988; Arthur, 1999; Beinhocker, 2006; Farmer, 2002; Miller & Page, 2007). The answer is they are close cousins but with some differences of emphasis. First, complex adaptive systems are generally thought of as multi-agent systems, but it is possible to imagine a reflexive system with one agent. Second, as noted, in reflexive systems internal model updating often involves not only changes in model parameters or weights, but changes in rules and model structure as well. Systems where agents have fixed rules but simply adjust rule parameters or weights in response to environmental feedback are often considered adaptive, but I would claim they are not necessarily reflexive in Soros's use of the term.

Thus, there is a high but imperfect degree of overlap between reflexive systems and complex adaptive systems. But given their close relationship, one might reasonably talk of 'complex reflexive systems' as a specific subset of the more general class of 'complex adaptive systems' which in turn is a subclass of the more general 'complex systems.'

4. Implications for social science

Space does not allow a full exploration of the many implications of Soros's ideas for epistemology, the philosophy of social science, and the future of economics. However, in this section, I will address one aspect of Soros's views: his arguments about the distinctions between natural and social science.

4.1 Timeless laws and prediction

Soros (2013, pp. 315–320) argues that one cannot create scientific explanations (at least in a Popperian sense) of human social systems because of the deep uncertainty and indeterminacy that reflexivity and fallibility introduce. Reflexive systems are inherently contingent and time bound, they do not have timeless universal laws, and one cannot make reliable predictions in such systems.

It is important to note, however, that all science is contingent and time bound – no science, not even physics has truly timeless laws. Chemistry has very empirically successful explanations and models of atomic interactions. Yet, atoms did not exist until 379,000 years after the Big Bang, so there was a period when the laws of chemistry did not exist. The regularities that chemistry describes are limited to a particular time and set of conditions. Likewise, the laws of evolution did not come into play until life was formed (whether on Earth or another planet). Some physicists even argue that time itself is not a fundamental property of the universe, but emerged as a post-Big Bang phenomenon too.

Rather than timeless, universal laws, what natural science describes is a process of the universe unfolding and creating itself after its origin, and new regularities and properties emerging at higher and higher levels of aggregation and complexity as that process proceeds. Those regularities may be universal for a period of time which may be very long, but any such regularities are nonetheless time bound and contingent. Seen from that perspective, both biological and social phenomena are merely at the end of a very long chain of growing complexity. So, the goal of science is not the discovery of timeless, universal laws – they don't exist. Rather, it is to explain time bound and contingent regularities in an empirically testable way.

Economics too has many regularities in need of explanation, for example, the power laws observed in financial volatility, Pareto distributions of income, the many regularities observed by behavioral economists, or the patterns of financial bubbles that Soros (2009, 2010, 2012) provides many insights into. These regularities in economic systems are dependent on a host of conditions and will not last forever – but neither will the laws of chemistry or biology and possibly even the laws of physics. The key is that the regularities are stable enough that they lend themselves to empirically testable explanation. Economics has many such stable regularities.

Soros (2013, p. 321) further argues that reflexivity destroys the ability to make 'determinate predictions' in social systems. Here, we must distinguish between prediction versus forecasting. Forecasting is a statement about the state of the universe at some future point in time. If Soros's claim is about forecasting then he is correct. As discussed in Section 1, reflexivity introduces indeterminacy and limits to knowledge that makes accurate forecasting in social systems often very difficult or even impossible.

But prediction, as it is used in a scientific sense, is a deductive logical consequence of a theory. A prediction does not necessarily have to take place in the future – the observation being predicted could take place now or in the past. For example, evolutionary theory cannot make accurate forecasts about the future evolution of individual species. The evolutionary process is too complex and suffers many of the same limits to knowledge issues characterized by reflexivity. But evolutionary theory has many testable, deductive consequences. For example, in 1859, Darwin wrote in the *Origin of Species* (Chapter 10) that his theory predicted fossils in pre-Cambrian rocks and if they were not found that it would be a strong argument against his theory – a kind of Popperian falsification test. In the 1950s and 1960s, with the help of more powerful microscopes such fossils were eventually found (Schopf, 2000). This was a prediction about past events, not a forecast of the future. Modern genetics has also validated many of the deductive implications of evolutionary theory in ways that Darwin himself could not have imagined.

Likewise, we may not be able to accurately predict future states of the stock market (though Mr Soros has an excellent record in this regard). But as noted there is a strong empirical regularity that price fluctuations are distributed according to a power law (e.g. Farmer & Lillo, 2004). Any theory of stock prices whose logical consequences included a distribution other than a power law would fail the Popperian test. Indeed, the efficient market hypothesis that predicts a random walk fails exactly this test.

Soros himself makes a number of testable predictions in his writings. For example, he has described in detail how reflexivity plays out in the laboratory of financial markets and how it underlies the dynamics of bubbles and crashes. Bubbles and crashes are an example of a persistent pattern or regularity. As Soros notes, bubbles cannot be forecast. But he shows how they can be explained and more deeply understood by this theory, and his theory predicts specific phases to bubbles that appear born out in the historical record. The predictions from reflexivity may be more qualitative than quantitative, but many natural sciences including evolutionary theory use descriptive and qualitative predictions very successfully.

4.2 A spectrum of complexity

Although the time frames of the regularities economics studies may be shorter than those in many physical systems, and forecasting may be difficult, neither of these disqualifies economics from potentially being a science. Instead, one can think of a spectrum of system complexity, stretching from basic mechanical systems (e.g. a ball rolling down a frictionless plane in a vacuum) to simple statistical systems (e.g. an ideal gas), to nonlinear, dynamic systems (e.g. a turbulent fluid), to systems with complex interactions (e.g. climate change), to complex adaptive systems (e.g. the human brain). Not only can physical systems be arrayed against such a spectrum, but so too can artificial and human social systems. Artificial systems range from the simple and mechanical (e.g. a pendulum swinging in a clock) to the mind-boggling complex (e.g. the Internet). Likewise, human social systems can be thought of along this spectrum – economics studies phenomena ranging from two people haggling over the price of a rug to the full complexity of the global economy.

Complex reflexive systems can be thought of as at the far end of this spectrum (on the right if one visualizes it stretching from left to right, see Figure 1). But a key question is whether reflexivity is limited to just human systems. Soros (2013, p. 311) defines reflexive systems as having 'thinking participants.' Numerous cognitive scientists and philosophers have felled many trees writing papers debating just what constitutes 'thinking' and whether thinking is something only humans can do. As Soros admits 'thinking' is hard to define and this creates a certain ambiguity in his definition. In my definition in Section 1, I have only generically referred to 'agents' whose models update, not specifically to humans. A bacterium swimming in a sugar gradient has a goal, a cognitive function, a manipulative function, an updating internal model, and two-way feedback with its

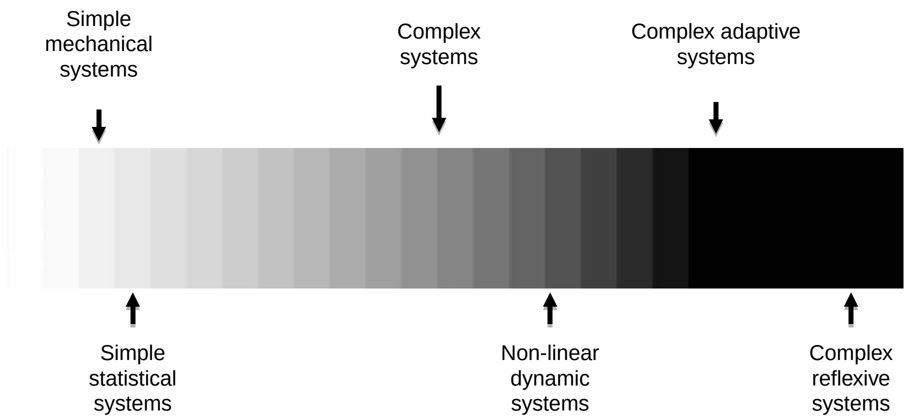

Figure 1. A spectrum of complexity.

environment. Under my definition, it would be a single-agent reflexive system. An evolving forest ecosystem would be a multi-agent complex reflexive system. It is also possible for artificial systems to be reflexive. For example, the black-box stock trading robots now used by many hedge funds have arguably created an artificial reflexive market.

Of course, Soros is correct in asserting that there is something special about human systems. The internal mental models of humans are far more complex than other species or any current artificial system, and humans are the only species to use language. Human social structures and institutions are also more complex and enduring than those of other species. But again it is a matter of degrees, sliding up the spectrum from bacterium, to lizard, to primate, to human. Rather than a black and white distinction one can think of human reflexive systems as the far, far right end of the spectrum of complexity.

4.3 Common epistemological challenges

If one accepts the spectrum of complexity argument then it has important implications for the nature of social science. What defines the epistemological challenge of understanding a particular phenomenon is *where it sits on the spectrum of complexity, not its domain.* Understanding and explaining two people playing a simple game theory problem with an easily calculated unique Nash equilibrium has more in common with a simple mechanical equilibrium system than it does with trying to understand the effect of contagion in a banking crisis. But understanding the effect of contagion in a banking crisis has some striking similarities to understanding contagion in epidemiology, or the collapse of a food web in ecology (Haldane & May, 2011).

At the left (simple) end of the spectrum of complexity the epistemological challenges are few. There is a clear regularity to be explained, data are available, there are few nonlinearities, dynamics, or interactions, and the problem is amenable to standard analytical, statistical, or computational methods. The possibility of accurate forecasting (not just prediction) for such systems is high. At the right (complex) end of the spectrum the situation is reversed. The epistemological challenges are many and include the limits to knowledge issues outlined in Section 1. And while Popperian prediction and testing may be possible, highly accurate forecasting may not be.

Soros is thus correct that reflexive systems are very challenging to model and understand scientifically. I would argue that they are challenging *not* because there is a fundamental difference between physical and social systems, but because they are extremely complex – whether it is an ecosystem of human beings trading in a stock market, an ecosystem of sophisticated stock trading computer algorithms, or an ecosystem of interacting species (Farmer, 2002).

4.4 Evolution, good enough models, and muddling through

The epistemological challenges may be high for understanding reflexive systems, but fallible models, limits to knowledge, and the inherent indeterminacy of reflexive systems do not necessarily imply that all is lost and that we cannot gain insight into such systems. Biological systems, human systems, and various artificial systems all manage to function despite fallibility and limits to knowledge. The internal models of agents and their cognitive and manipulative functions may be 'good enough' for them to muddle through and make progress toward their goals. As the statistician George E.P. Box once said, 'Essentially, all models are wrong, but some are useful' (Box, 1987, p. 424).

Indeed, biological evolution depends on 'useful enough' models muddling through without the ability to forecast. Evolution picks up on regularities and through experimentation, selection, and amplification finds heuristics that are 'good enough for now' until something better comes along or something selects against them. There is a parallel with Popper's (1972, 1984) evolutionary views of science and epistemology, and in particular his view on the role of falsification in science. Success for a theory is always highly contingent and subject to future refutation while failure is generally fatal. This asymmetry is true for any evolutionary epistemological system where objective truth is inaccessible but increasingly successful approximations are possible. This has been the core idea of the evolutionary epistemology of theories program in the philosophy of science (Bradie, 1986; Campbell, 1960; Hull, 1988; Toulmin, 1972). And in such systems monotonic progress is not guaranteed – biological evolution, human social systems, and science can all experience blind alleys, reversals, and collapses. A model is shown to be wrong, and the search is on again for new and better ones. I would even claim that reflexive systems *require* evolutionary epistemological processes to deal with fallibility and limits to knowledge – there is simply no other way for such systems to function in a noisy, complex, difficult to forecast world.

I should also briefly note that while complex reflexive systems present a challenge due to their Knightian uncertainty, there is also an upside – their inherent indeterminacy creates space for novelty and creativity. In a perfectly deterministic system, there is no room for novelty to emerge. In a perfectly random system, there is insufficient coherence for novelty to matter. Complex reflexive systems are somewhere in between – they are hard to forecast and fallibility is inevitable, but there is nonetheless pattern and structure. This creates a space for novelty, experimentation, and in human systems creativity. One may even argue that there is a link between reflexivity and free will – reflexivity makes free will both possible and necessary.

4.5 Model-dependent realism: reconciling Soros and Popper

Reflexivity is not only consistent with an evolutionary view of epistemology, but also what the physicists Stephen Hawking and Leonard Mlodinow call 'model-dependent realism' (Hawking & Mlodinow, 2010). Under this view, there may or may not be an objective reality independent from us and the models we create. But whether there is or not does not matter because the only way we can access and perceive our world is via the models we create – whether it is the model of the room I am sitting in that my brain assembles from pixels of light coming through my eyes, or a mathematical physics model of an atom. One cannot separate reality from our models or say in any sense that a model is objectively true – all one can say is whether a particular model fits the data one has better than another model (in which case it is 'the best we have for now') or whether the data falsify the model. Such a pragmatic stance is an attempt to resolve the age-old debate in philosophy of science between 'realists' who argue that science is about discovering fundamental truths and explaining phenomena, and 'instrumentalists' who argue for more the more limited goal of creating tools that can make reliable and useful predictions.

If one combines evolutionary epistemology with model-dependent realism one comes to a view that science is about providing explanations of phenomena (realism), but such explanations are always mediated by our models and observations and thus cannot claim to be objectively and perpetually true (model dependency). Scientific explanations are formulated as models or theories which have deductive implications, some of which are observable and some are not (a model with no observable deductive implications is not scientific). Observable implications can then be used to make testable predictions. If those predictions are more congruent with empirical observation versus competing explanations, then the theory or model is 'the best explanation for now' (empirical adequacy), if it is falsified then it is thrown out (Popperian selection). Science is thus an evolutionary epistemological system of competing explanations, with the primary selection force being Popperian falsifiability (although, as many critics of Popper have argued, in real science falsifiability is rarely black and white and thus theory selection is often a complex mix of criteria – some empirically based and some more sociological, e.g. the power of authority figures in a field). The evolutionary process of competing explanations over time leads to the paradigmatic eras described by Kuhn (1962).

I believe that such an evolutionary, model-dependent realist epistemology is largely consistent with Soros's (2013) views. A passage from Hawking and Mlodinow (2010, Chap. 3) sounds almost as if it could have been written by Soros:

> Model-dependent realism applies not only to scientific models but also to the conscious and subconscious mental models we all create in order to interpret and understand the everyday world. There is no way to remove the observer – us – from our perception of the world, which is created through our sensory processing and through the way we think and reason.

As Soros (2013, pp. 316–317) observes, in reflexive systems agents do not have direct access to any objective reality – there is no ideal of the detached, objective observer. I would claim that in reflexive systems the best agents can do is to be evolutionary model-dependent realists and make judgments, using Popper's great insight of falsifiability, as to whether one model fits the finite, noisy, flawed data they observe better than competing models. In this way, inherently fallible agents muddle through with inherently imperfect models. Yet, these models may improve over time, even if they are only for a time, and in this way our knowledge grows.

5. A way forward for economics

Economies are certainly reflexive system and at the far end of the spectrum of complexity. Soros is correct in saying that the reflexivity of economic systems has been largely ignored by mainstream economics. For the past 130 years since Walras, economists have insisted on treating economies as simple mechanical or statistical systems. Economics is a diverse subject, but the core theories that have dominated the field and policy-making have been equilibrium theories derived deductively from axiomatic principles (Hausman, 1992).

However, much empirical work, particularly in behavioral economics, has shown that many of the core axiomatic assumptions of economic theory are flawed. Economics has what computer scientists call a 'garbage in, garbage out' problem – the logic of its proofs and equations may be flawless – but it is built on a base of empirically disproven assumptions and so the conclusions based on those assumptions are likely to be suspect as well (Beinhocker, 2006). This is born out by meta studies of empirical results in economics that conclude that the empirical credibility of economics is 'modest or even low' (Ioannides & Doucouliagos, 2013, p. 1).

To move forward and develop into what truly could be called a science, economics needs to embrace the philosophy of both Soros and Popper – as well as the toolkit of the modern science of complex systems. Economics needs to recognize that it made an ontological error when in the nineteenth century it categorized economies as equilibrium systems (Beinhocker, 2006; Mirowski, 1989). It was perhaps an understandable error as equilibrium analysis was the tool available at the time – but now we know better. Following Soros and categorizing economies as complex reflexive systems would end the false certainty of neoclassical theory and enable economists to embrace the inherent fallibility and Knightian uncertainty that characterizes real-world economic systems.

Soros (2013, p. 317) argues that economics has made an error in trying to be too much like the other sciences, but I would argue it has made too little effort. Economics has not been Popperian enough. It has developed an axiomatic, internally consistent, self-contained theory – more like theology than a science – and has developed a culture that is resistant to empirical testing and falsification (Ioannides & Doucouliagos, 2013). Theories such as rational expectations and the efficient markets hypothesis have been thoroughly discredited by decades of empirical work in behavioral, experimental, statistical, and financial economics – not to mention the real-world experiences of the 2008 financial crisis – and yet remain in current use in the field and in policy.

Economists must give up the myth that the economy is a simple mechanical equilibrium system. The field needs to embrace the economy in its full messy, uncertain, disequilibrium, complex reflexive reality. Approaches that can help give insight into such a system range from those found in the behavioral sciences, to experimental economics, network theory, nonlinear systems theory, complexity theory, evolutionary theory, and information theory. Economics is an inherently multi-disciplinary field and should draw ideas from across the social and natural sciences, as well as utilize a plurality of methods from analytical, to statistical, to computational, to qualitative and historical.

Soros has presented us with a major epistemological challenge, one that economics has not properly grappled with to date. Economics will never look like the Newtonian physics that Walras and many since have dreamed of. But neither does climate change science, neuroscience, evolutionary theory, or any other scientific field that studies highly complex phenomena. Human, complex reflexive systems may be another level more complex than these. Yet, we can be hopeful that although our ability to understand such systems may always be limited, our creativity in trying to will not be.

Acknowledgements

The author is grateful for the support of the Institute for New Economic Thinking (INET) and thanks the participants of the workshops on reflexivity at Central European University, Budapest, November 2011 and October 2013 for their comments and suggestions. The author also benefited enormously from George Soros's insights and his Popperian 'critical attitude.' All errors are the author's.

References

Anderson, P. W., Arrow, K. J., & Pines, D. (Eds.). (1988). *The economy as a complex evolving system*. Redwood City, CA: Addison-Wesley.

Arthur, W. B. (1999). Complexity and the economy. *Science*, *284*, 107–109.

Beinhocker, E. D. (2006). *The origin of wealth: Evolution, complexity and the radical remaking of economics*. Boston, MA: Harvard Business School Press.

Box, G. E. P., & Draper, N. R. (1987). *Empirical model building and response surfaces*. New York, NY: Wiley.

Bradie, M. (1986). Assessing evolutionary epistemology. *Biology and Philosophy*, *1*, 401–459.

Campbell, D. T. (1960). Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review*, *67*, 380–400.

Farmer, J. D. (2002). Market force, ecology, and evolution. *Industrial and Corporate Change*, *11*, 895–953.

Farmer, J. D., & Lillo, F. (2004). On the origin of power law tails in price fluctuations. *Journal of Quantitative Finance*, *4*, C7–C11.

Forrester, J. (1961). *Industrial dynamics*. Waltham, MA: Pegasus.

Haldane, A. G., & May, R. M. (2011). Systemic risk in banking ecosystems. *Nature*, *469*, 351–355.

Hausman, D. M. (1992). *The inexact and separate science of economics*. Cambridge: Cambridge University Press.

Hawking, S., & Mlodinow, L. (2010). *The grand design*. New York, NY: Bantam Books.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.

Hommes, C. (2013). Reflexivity, expectations feedback and almost self-fulfilling equilibrium: Economic theory, empirical evidence and laboratory experiments. *Journal of Economic Methodology*, *20*, 406–419.

Hull, D. L. (1988). *Science as a process: An evolutionary account of the social and conceptual development of science*. Chicago, IL: University of Chicago Press.

Ioannides, J., & Doucouliagos, C. (2013). What's to know about the credibility of empirical economics? *Journal of Economic Surveys*, *27*, 997–1004.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Lucas, R. E., & Prescott, E. C. (1971). Investment under uncertainty. *Econometrica*, *39*, 659–681.

Miller, J. H., & Page, S. E. (2007). *Complex adaptive systems: An introduction to computational models of social life*. Princeton, NJ: Princeton University Press.

Mirowski, P. (1989). *More heat than light: Economics as social physics, physics as nature's economics*. Cambridge: Cambridge University Press.

Muth, J. E. (1961). Rational expectations and the theory of price movements. *Econometrica*, *29*, 315–335.

Popper, K. R. (1935). *Logik der Forschung*. Vienna: First English edition (1959) *The logic of scientific discovery*, London: Hutchinson Verlag von Julius Springer.

Popper, K. R. (1972). *Objective knowledge: An evolutionary approach*. Oxford: Clarendon Press.

Popper, K. R. (1984). Evolutionary epistemology. In J. W. Pollard (Ed.), *Evolutionary theory: Paths into the future* (pp. 239–254). London: Wiley.

Schopf, J. W. (2000). Solution to Darwin's dilemma: Discovery of the missing Precambrian record of life. *Proceedings of the National Academy of Sciences*, *97*, 6947–6953.

Soros, G. (1987). *The Alchemy of finance*. Hoboken, NJ: Wiley.

Soros, G. (1998). *The crisis of global capitalism: Open society endangered*. New York, NY: Public Affairs.

Soros, G. (2000). *Open society: Reforming global capitalism*. New York, NY: Public Affairs.

Soros, G. (2006). *The age of fallibility: Consequences of the war on terror*. New York, NY: Public Affairs.

Soros, G. (2008). *The new paradigm for financial markets the credit crisis and what it means*. New York, NY: Public Affairs.

Soros, G. (2009). *The crash of 2008 and what it means: The new paradigm for financial markets*. New York, NY: Public Affairs.

Soros, G. (2010). *The Soros lectures at the Central European University*. New York, NY: Public Affairs.

Soros, G. (2012). *Financial Turmoil in the United States and Europe: Essays*. New York, NY: Public Affairs.

Soros, G. (2013). Fallibility, reflexivity and the human uncertainty principle. *Journal of Economic Methodology*, *20*, 309–329.
Sterman, J. (2000). *Business dynamics: Systems thinking and modelling for a complex world*. New York, NY: McGraw Hill.
Toulmin, S. (1972). *Human understanding: The collective use and evolution of concepts*. Princeton, NJ: Princeton University Press.
Umpleby, S. (2007). Reflexivity in social systems: The theories of George Soros. *Systems Research and Behavioral Science*, *24*, 515–522.
Umpleby, S. (2010). From complexity to reflexivity: Underlying logics used in science. *Journal of the Washington Academy of Sciences*, *96*, 15–26.
Wiener, N. (1948). *Cybernetics or control and communication in the animal and machine*. New York, NY: Wiley.