

# Transcript: General Theory of Reflexivity

*The following is a transcript of a lecture given by George Soros at Central European University on October 26, 2009.*

In the course of my life, I have developed a conceptual framework which has helped me both to make money as a hedge fund manager and to spend money as a policy oriented philanthropist. But the framework itself is not about money, it is about the relationship between thinking and reality, a subject that has been extensively studied by philosophers from early on.

I started developing my philosophy as a student at the London School of Economics in the late 1950s. I took my final exams one year early and I had a year to fill before I was qualified to receive my degree. I could choose my tutor and I chose Karl Popper, the Viennese-born philosopher whose book *The Open Society and Its Enemies* had made a profound impression on me.

In his books Popper argued that the empirical truth cannot be known with absolute certainty. Even scientific laws can't be verified beyond a shadow of a doubt: they can only be falsified by testing. One failed test is enough to falsify, but no amount of conforming instances is sufficient to verify. Scientific laws are hypothetical in character and their truth remains subject to testing. Ideologies which claim to be in possession of the ultimate truth are making a false claim; therefore, they can be imposed on society only by force. This applies to Communism, Fascism and National Socialism alike. All these ideologies lead to repression. Popper proposed a more attractive form of social organization: an open society in which people are free to hold divergent opinions and the rule of law allows people with different views and interests to live together in peace. Having lived through both Nazi and Communist occupation here in Hungary I found the idea of an open society immensely attractive.

While I was reading Popper I was also studying economic theory and I was struck by the contradiction between Popper's emphasis on imperfect understanding and the theory of perfect competition in economics which postulated perfect knowledge. This led me to start questioning the *assumptions* of economic theory. These were the two major *theoretical* inspirations of my philosophy. It is also deeply rooted in my personal history.

The formative experience of my life was the German occupation of Hungary in 1944. I was not yet fourteen years old at the time, coming from a reasonably well-to-do middle class background, suddenly confronted with the prospect of being deported and killed just because I was Jewish.

Fortunately my father was well prepared for this far-from-equilibrium experience. He had lived through the Russian Revolution and that was the formative experience of *his* life. Until then he had been an ambitious young man. When the First World War broke out, he volunteered to serve in the Austro-Hungarian army. He was captured by the Russians and taken as a prisoner of war to Siberia. Being ambitious, he became the editor of a newspaper produced by the prisoners. It

was handwritten and displayed on a plank and it was called *The Plank*. This made him so popular that he was elected the prisoners' representative. Then some soldiers escaped from a neighboring camp, and their prisoners' representative was shot in retaliation. My father, instead of waiting for the same thing to happen in his camp, organized a group and led a breakout. His plan was to build a raft and sail down to the ocean, but his knowledge of geography was deficient; he did not know that all the rivers in Siberia flow into the Arctic Sea. They drifted for several weeks before they realized that they were heading for the Arctic, and it took them several more months to make their way back to civilization across the taiga. In the meantime, the Russian Revolution broke out, and they became caught up in it. Only after a variety of adventures did my father manage to find his way back to Hungary; had he remained in the camp, he would have arrived home much sooner.

My father came home a changed man. His experiences during the Russian Revolution profoundly affected him. He lost his ambition and wanted nothing more from life than to enjoy it. He imparted to his children values that were very different from those of the milieu in which we lived. He had no desire to amass wealth or become socially prominent. On the contrary, he worked only as much as was necessary to make ends meet. I remember being sent to his main client to borrow some money before we went on a ski vacation; my father was grouchy for weeks afterwards because he had to work to pay it back. Although we were reasonably prosperous, we were not the typical bourgeois family, and we were proud of being different.

In 1944, when the Germans occupied Hungary, my father immediately realized that these were not normal times and the normal rules didn't apply. He arranged false identities for his family and a number of other people. Those who could, paid; others he helped for free. Most of them survived. That was his finest hour.

\* \* \*

Living with false identity turned out to be an exhilarating experience for me too. We were in mortal danger. People perished all around us, but we managed not only to survive but to help other people. We were on the side of the angels, and we triumphed against overwhelming odds. This made me feel very special. It was high adventure. I had a reliable guide in my father and came through unscathed. What more could a fourteen-year-old ask for?

After the euphoric experience of escaping the Nazis, life in Hungary started to lose its luster during the Soviet occupation. I was looking for new challenges and with my father's help I found my way out of Hungary. When I was seventeen I became a student in London. In my studies, my primary interest was to gain a better understanding of the strange world into which I had been born, but I have to confess, I also harbored some fantasies of becoming an important philosopher. I believed that I had gained insights that set me apart from other people.

Living in London was a big letdown. I was without money, alone, and people were not interested in what I had to say. But I didn't abandon my philosophical ambitions even when circumstances forced me to make a living in more mundane pursuits. After completing my studies, I had a

number of false starts. Finally I ended up as an arbitrage trader in New York but in my free time I continued to work on my philosophy.

That is how I came to write my first major essay, entitled "The Burden of Consciousness." It was an attempt to model Popper's framework of open and closed societies. It linked organic society with a traditional mode of thinking, closed society with a dogmatic mode and open society with a critical mode. What I could not properly resolve was the nature of the relationship between the mode of thinking and the actual state of affairs. That problem continued to preoccupy me and that is how I came to develop the concept of reflexivity—a concept I shall explore in greater detail a little later.

It so happened that the concept of reflexivity provided me with a new way of looking at financial markets, a better way than the prevailing theory. This gave me an edge, first as a securities analyst and then as a hedge fund manager. I felt as if I were in possession of a major discovery that would enable me to fulfill my fantasy of becoming an important philosopher. At a certain moment when my business career ran into a roadblock I shifted gears and devoted all my energies to developing my philosophy. But I treasured my discovery so much that I could *not* part with it. I felt that the concept of reflexivity needed to be explored in depth. As I delved deeper and deeper into the subject I got lost in the intricacies of my own constructions. One morning I could not understand what I had written the night before. At that point I decided to abandon my philosophical explorations and to focus on making money. It was only many years later, after a successful run as a hedge fund manager, that I returned to my philosophy.

I published my first book, *The Alchemy of Finance*, in 1987. In that book I tried to explain the philosophical underpinnings of my approach to financial markets. The book attracted a certain amount of attention. It has been read by most people in the hedge fund industry and it is taught in business schools but the philosophical arguments did not make much of an impression. They were largely dismissed as the conceit of a man who has been successful in business and fancied himself as a philosopher.

I myself came to doubt whether I was in possession of a major new insight. After all I was dealing with a subject that has been explored by philosophers since time immemorial. What grounds did I have for thinking that I had made a new discovery, especially as nobody else seemed to think so? Undoubtedly the conceptual framework was useful to me personally but it did not seem to be considered equally valuable by others. I had to accept their judgment. I didn't give up my philosophical interests, but I came to regard them as a personal predilection. I continued to be guided by my conceptual framework both in my business and in my philanthropic activities—which came to assume an increasingly important role in my life—and each time I wrote a book I faithfully recited my arguments. This helped me to develop my conceptual framework, but I continued to consider myself a failed philosopher. Once I even gave a lecture with the title "A Failed Philosopher Tries Again."

All this has changed as a result of the financial crisis of 2008. My conceptual framework enabled me both to anticipate the crisis and to deal with it when it finally struck. It has also enabled me to

explain and predict events better than most others. This has changed my own evaluation and that of many others. My philosophy is no longer a personal matter; it deserves to be taken seriously as a possible contribution to our understanding of reality. That is what has prompted me to give this series of lectures.

\* \* \*

So here it goes. Today I shall explain the concepts of fallibility and reflexivity in general terms. Tomorrow I shall apply them to the financial markets and after that, to politics. That will also bring in the concept of open society. In the fourth lecture I shall explore the difference between market values and moral values, and in the fifth I shall offer some predictions and prescriptions for the present moment in history.

\* \* \*

I can state the core idea in two relatively simple propositions. One is that in situations that have thinking participants, the participants' view of the world is always partial and distorted. That is the principle of fallibility. The other is that these distorted views can influence the situation to which they relate because false views lead to inappropriate actions. That is the principle of reflexivity. For instance, treating drug addicts as criminals creates criminal behavior. It misconstrues the problem and interferes with the proper treatment of addicts. As another example, declaring that government is bad tends to make for bad government.

Both fallibility and reflexivity are sheer common sense. So when my critics say that I am merely stating the obvious, they are right-but only up to a point. What makes my propositions interesting is that their significance has not been generally appreciated. The concept of reflexivity, in particular, has been studiously avoided and even denied by economic theory. So my conceptual framework deserves to be taken seriously-not because it constitutes a new discovery but because something as commonsensical as reflexivity has been so studiously ignored.

Recognizing reflexivity has been sacrificed to the vain pursuit of certainty in human affairs, most notably in economics, and yet, uncertainty is the key feature of human affairs. Economic theory is built on the concept of equilibrium, and that concept is in direct contradiction with the concept of reflexivity. As I shall show in the next lecture, the two concepts yield two entirely different interpretations of financial markets.

The concept of fallibility is far less controversial. It is generally recognized that the complexity of the world in which we live exceeds our capacity to comprehend it. I have no great new insights to offer. The main source of difficulties is that participants are part of the situation they have to deal with. Confronted by a reality of extreme complexity we are obliged to resort to various methods of simplification-generalizations, dichotomies, metaphors, decision-rules, moral precepts, to mention just a few. These mental constructs take on an existence of their own, further complicating the situation.

The structure of the brain is another source of distortions. Recent advances in brain science have begun to provide some insight into how the brain functions, and they have substantiated Hume's contention that reason is the slave of passion. The idea of a disembodied intellect or reason is a figment of our imagination.

The brain is bombarded by millions of sensory impulses but consciousness can process only seven or eight subjects concurrently. The impulses need to be condensed, ordered and interpreted under immense time pressure, and mistakes and distortions can't be avoided. Brain science adds many new details to my original contention that our understanding of the world in which we live is inherently imperfect.

\* \* \*

The concept of reflexivity needs a little more explication. It applies exclusively to situations that have thinking participants. The participants' thinking serves two functions. One is to understand the world in which we live; I call this the cognitive function. The other is to change the situation to our advantage. I call this the participating or manipulative function. The two functions connect thinking and reality in opposite directions. In the cognitive function, reality is supposed to determine the participants' views; the direction of causation is from the world to the mind. By contrast, in the manipulative function, the direction of causation is from the mind to the world, that is to say, the intentions of the participants have an effect on the world. When both functions operate at the same time they can interfere with each other.

How? By depriving each function of the independent variable that would be needed to determine the value of the dependent variable. Because, when the independent variable of one function is the dependent variable of the other, neither function has a genuinely independent variable. This means that the cognitive function can't produce enough knowledge to serve as the basis of the participants' decisions. Similarly, the manipulative function can have an effect on the outcome, but can't determine it. In other words, the outcome is liable to diverge from the participants' intentions. There is bound to be some slippage between intentions and actions and further slippage between actions and outcomes. As a result, there is an element of uncertainty both in our understanding of reality and in the actual course of events.

To understand the uncertainties associated with reflexivity, we need to probe a little further. If the cognitive function operated in isolation without any interference from the manipulative function it could produce knowledge. Knowledge is represented by true statements. A statement is true if it corresponds to the facts-that is what the correspondence theory of truth tells us. But if there is interference from the manipulative function, the facts no longer serve as an independent criterion by which the truth of a statement can be judged because the correspondence may have been brought about by the statement changing the facts.

Consider the statement, "it is raining." That statement is true or false depending on whether it is, in fact, raining. Now consider the statement, "This is a revolutionary moment." That statement is reflexive, and its truth value depends on the impact it makes.

Reflexive statements have some affinity with the paradox of the liar, which is a self-referential statement. But while self-reference has been extensively analyzed, reflexivity has received much less attention. This is strange, because reflexivity has an impact on the real world, while self-reference is purely a linguistic phenomenon.

In the real world, the participants' thinking finds expression not only in statements but also, of course, in various forms of action and behavior. That makes reflexivity a very broad phenomenon that typically takes the form of feedback loops. The participants' views influence the course of events, and the course of events influences the participants' views. The influence is continuous and circular; that is what turns it into a feedback loop.

Reflexive feedback loops have *not* been rigorously analyzed and when I originally encountered them and tried to analyze them, I ran into various complications. The feedback loop is supposed to be a two-way connection between the participant's views and the actual course of events. But what about a two-way connection between the participants' views? And what about a solitary individual asking himself who he is and what he stands for and changing his behavior as a result of his reflections? In trying to resolve these difficulties I got so lost among the categories I created that one morning I couldn't understand what I had written the night before. That's when I gave up philosophy and devoted my efforts to making money.

To avoid that trap let me propose the following terminology. Let us distinguish between the objective and subjective aspects of reality. Thinking constitutes the subjective aspect, events the objective aspect. In other words, the subjective aspect covers what takes place in the minds of the participants, the objective aspect denotes what takes place in external reality. There is only one external reality but many different subjective views. Reflexivity can then connect any two or more aspects of reality, setting up two-way feedback loops between them. Exceptionally it may even occur with a single aspect of reality, as in the case of a solitary individual reflecting on his own identity. This may be described as "self-reflexivity." We may then distinguish between two broad categories: reflexive *relationships* which connect the subjective aspects and reflexive *events* which involve the objective aspect. Marriage is a reflexive relationship; the Crash of 2008 was a reflexive event. When reality has no subjective aspect, there can be no reflexivity.

\* \* \*

Feedback loops can be either negative or positive. Negative feedback brings the participants' views and the actual situation closer together; positive feedback drives them further apart. In other words, a negative feedback process is self-correcting. It can go on forever and if there are no significant changes in external reality, it may eventually lead to an *equilibrium* where the participants' views come to correspond to the actual state of affairs. That is what is supposed to happen in financial markets. So equilibrium, which is the *central* case in economics, turns out to be an extreme case of negative feedback, a *limiting* case in my conceptual framework.

By contrast, a positive feedback process is self-reinforcing. It cannot go on forever because eventually the participants' views would become so far removed from objective reality that the



participants would have to recognize them as unrealistic. Nor can the iterative process occur without any change in the actual state of affairs, because it is in the nature of positive feedback that it reinforces whatever tendency prevails in the real world. Instead of equilibrium, we are faced with a dynamic disequilibrium or what may be described as far-from-equilibrium conditions. Usually in far-from-equilibrium situations the divergence between perceptions and reality leads to a climax which sets in motion a positive feedback process in the *opposite* direction. Such initially self-reinforcing but eventually self-defeating boom-bust processes or bubbles are characteristic of financial markets, but they can also be found in other spheres. There, I call them fertile fallacies-interpretations of reality that are distorted, yet produce results which reinforce the distortion.

\* \* \*

I realize that this is all very abstract and difficult to follow. It would make it much easier if I gave some concrete examples. But you will have to bear with me. I want to make a different point and the fact that it is difficult to follow abstract arguments helps me make it. In dealing with subjects like reality or thinking or the relationship between the two, it's easy to get confused and formulate problems the wrong way. So misinterpretations and misconceptions can play a very important role in human affairs. The recent financial crisis can be attributed to a mistaken interpretation of how financial markets work. I shall discuss that in the next lecture. In the third lecture, I shall discuss two fertile fallacies-the Enlightenment fallacy and the post-modern fallacy. These concrete examples will demonstrate how important misconceptions have been in the course of history. But for the rest of this lecture I shall stay at the lofty heights of abstractions.

I contend that situations that have thinking participants have a different structure from natural phenomena. The difference lies in the role of thinking. In natural phenomena thinking plays no *causal role* and serves only a cognitive function. In human affairs thinking is *part* of the subject matter and serves both a cognitive and a manipulative function. The two functions can interfere with each other. The interference does not occur all the time-in everyday activities, like driving a car or painting a house, the two functions actually complement each other-but *when* it occurs, it introduces an element of uncertainty which is absent from natural phenomena. The uncertainty manifests itself in *both* functions: the participants' act on the basis of imperfect understanding and the results of their actions will not correspond to their expectations. That is a key feature of human affairs.

By contrast, in the case of natural phenomena, events unfold irrespective of the views held by the observers. The outside observer is engaged only in the cognitive function and the phenomena provide a reliable criterion by which the truth of the observers' theories can be judged. So the outside observer can obtain knowledge. Based on that knowledge, nature can be successfully manipulated. There is a natural separation between the cognitive and manipulative functions. Due to their separation, both functions can serve their purpose better than in the human sphere.

At this point, I need to emphasize that reflexivity is not the only source of uncertainty in human affairs. *Yes*, reflexivity does introduce an element of uncertainty both into the participants views and the actual course of events, but *other* factors may also have the same effect. For instance, the fact that participants cannot know what the other participants know, is something quite different from reflexivity, yet it is a source of uncertainty in human affairs. The fact that different participants have different interests, some of which may be in conflict with each other, is another source of uncertainty. Moreover, each individual participant may be guided by a multiplicity of values which may not be self-consistent, as Isaiah Berlin pointed out. The uncertainties created by these factors are likely to be even more extensive than those generated by reflexivity. I shall lump them all together and speak of the *human uncertainty principle*, which is an even broader concept than reflexivity.

The human uncertainty principle I am talking about is much more specific and stringent than the subjective skepticism that pervades Cartesian philosophy. It gives us objective reasons to believe that our perceptions and expectations are-or at least may be-wrong.

Although the primary impact of human uncertainty falls on the participants, it has far-reaching implications for the social sciences. I can explicate them best by invoking Karl Popper's theory of scientific method. It is a beautifully simple and elegant scheme. It consists of three elements and three operations. The three elements are scientific laws and the initial and final conditions to which those laws apply. The three operations are prediction, explanation, and testing. When the scientific laws are combined with the initial conditions, they provide predictions. When they are combined with the final conditions, they provide explanations. In this sense predictions and explanations are symmetrical and reversible. That leaves testing, where predictions derived from scientific laws are compared with the actual results.

According to Popper, scientific laws are hypothetical in character; they cannot be verified, but they can be falsified by testing. The key to the success of scientific method is that it can test generalizations of universal validity with the help of singular observations. One failed test is sufficient to falsify a theory but no amount of confirming instances is sufficient to verify.

This is a brilliant solution to the otherwise intractable problem: how can science be both empirical and rational? According to Popper it is empirical because we *test* our theories by observing whether the predictions we derive from them are true, and it is rational because we use deductive logic in doing so. Popper dispenses with inductive logic and relies instead on testing. Generalizations that cannot be falsified, do not qualify as scientific. Popper emphasizes the central role that testing plays in scientific method and establishes a strong case for critical thinking by asserting that scientific laws are only provisionally valid and remain open to reexamination. Thus the three salient features of Popper's scheme are the symmetry between prediction and explanation, the *asymmetry* between verification and falsification and the central role of testing. Testing allows science to grow, improve and innovate.

Popper's scheme works well for the study of natural phenomena but the human uncertainty principle throws a monkey wrench into the supreme simplicity and elegance of Popper's scheme.



The symmetry between prediction and explanation is destroyed because of the element of uncertainty in predictions and the central role of testing is endangered. Should the initial and final conditions include or exclude the participant's thinking? The question is important because testing requires replicating those conditions. If the participants' thinking is included, it is difficult to observe what the initial and final conditions are, because the participants' views can only be inferred from their statements or actions. If it is excluded, the initial and final conditions do not constitute singular observations because the same objective conditions may be associated with very different views held by the participants. In either case, generalizations cannot be properly tested. These difficulties do not preclude social scientists from producing worthwhile generalizations, but they are unlikely to meet the requirements of Popper's scheme, nor can they match the predictive power of the laws of physics.

Social scientists have found this conclusion hard to accept. Economists in particular suffer from what Sigmund Freud might call "physics envy."

There have been many attempts to eliminate the difficulties connected with the human uncertainty principle by inventing or postulating some kind of fixed relationship between the participants' thinking and the actual state of affairs. Karl Marx asserted that the ideological superstructure was determined by the material conditions of production and Freud maintained that people's behavior was determined by drives and complexes of which they were not even conscious. Both claimed scientific status for their theories although, as Popper pointed out, they cannot be falsified by testing.

But by far the most impressive attempt has been mounted by economic theory. It started out by assuming perfect knowledge and when that assumption turned out to be untenable it went through ever increasing contortions to maintain the fiction of rational behavior. Economics ended up with the theory of rational expectations which maintains that there is a single optimum view of the future, that which corresponds to it, and eventually all the market participants will converge around that view. This postulate is absurd but it is needed in order to allow economic theory to model itself on Newtonian physics.

Interestingly, both Karl Popper and Friedrich Hayek recognized, in their famous exchange in the pages of *Economica*, that the social sciences cannot produce results comparable to physics. Hayek inveighed against the mechanical and uncritical application of the quantitative methods of natural science. He called it scientism. And Karl Popper wrote about "The Poverty of Historicism" where he argued that history is not determined by universally valid scientific laws.

Nevertheless, Popper proclaimed what he called the "doctrine of the unity of method" by which he meant that both natural and social sciences should be judged by the same criteria. And Hayek, of course, became the apostle of the Chicago school of economics where market fundamentalism originated. But as I see it, the implication of the human uncertainty principle is that the subject matter of the natural and social sciences is fundamentally different; therefore they need to develop different methods and they have to be held to different standards. Economic theory should not be expected to produce universally valid laws that can be used reversibly to explain

and predict historic events. I contend that the slavish imitation of natural science inevitably leads to the distortion of human and social phenomena. What is attainable in social science falls short of what is attainable in physics.

I am somewhat troubled, however, about drawing too sharp a distinction between natural and social science. Such dichotomies are usually not found in reality; they are introduced by us, in our efforts to make some sense out of an otherwise confusing reality. Indeed while a sharp distinction between physics and social sciences seems justified, there are other sciences, such as biology and the study of animal societies that occupy intermediate positions.

But I had to abandon my reservations and recognize a dichotomy between the natural and social sciences because the social sciences encounter a second difficulty from which the natural sciences are exempt.

And that is that social theories are reflexive. Heisenberg's discovery of the uncertainty principle did not alter the behavior of quantum particles one iota, but social theories, whether Marxism, market fundamentalism or the theory of reflexivity, can affect the subject matter to which it refers. Scientific method is supposed to be devoted to the pursuit of truth. Heisenberg's uncertainty principle does not interfere with that postulate but the reflexivity of social theories does. Why should social science confine itself to passively studying social phenomena when it can be used to actively change the state of affairs? As I remarked in *The Alchemy of Finance*, the alchemists made a mistake in trying to change the nature of base metals by incantation. Instead, they should have focused their attention on the financial markets where they could have succeeded.

How could social science be protected against this interference? I propose a simple remedy: recognize a dichotomy between the natural and social sciences. This will ensure that social theories will be judged on their merits and not by a false analogy with natural science. I propose this as a convention for the protection of scientific method, not as a demotion or devaluation of social science. The convention sets no limits on what social science may be able to accomplish. On the contrary, by liberating social science from the slavish imitation of natural science and protecting it from being judged by the wrong standards, it should open up new vistas. It is in this spirit that I shall put forward my interpretation of financial markets tomorrow.

I apologize for dwelling so long in the rarefied realm of abstractions. I promise to come down to earth in my next lecture.