The Theory of Reflexivity By George Soros

Delivered April 26, 1994 to the MIT Department of Economics World Economy Laboratory Conference Washington, D.C.

When Rudi Dornbusch invited me to speak at this conference, he gave me a totally free hand in deciding what I wanted to talk about. Well, I want to discuss a subject which fascinates me but doesn't seem to interest others very much. That is my theory of reflexivity which has guided me both in making money and in giving money away, but has received very little serious consideration from anybody else. It is really a very curious situation. I am taken very seriously; indeed, a bit too seriously. But the theory that I take seriously and, in fact, rely on in my decision-making process is pretty completely ignored. I have written a book about it which was first published in 1987 under the title The Alchemy of Finance; but it received practically no critical examination. It has been out of print for the last several years but demand has been building up as a result of my increased visibility, not to say notoriety, and now the book is being re-issued. I think this is a good time to get the theory seriously considered.

I was invited to testify before Congress last week and this is how I started my testimony. I quote: "I must state at the outset that I am in fundamental disagreement with the prevailing wisdom. The generally accepted theory is that financial markets tend towards equilibrium, and on the whole, discount the future correctly. I operate using a different theory, according to which financial markets cannot possibly discount the future correctly because they do not merely discount the future; they help to shape it. In certain circumstances, financial markets can affect the so-called fundamentals which they are supposed to reflect. When that happens, markets enter into a state of dynamic disequilibrium and behave quite differently from what would be considered normal by the theory of efficient markets. Such boom/bust sequences do not arise very often, but when they do, they can be very disruptive, exactly because they affect the fundamentals of the economy." I did not have time to expound my theory before Congress, so I am taking advantage of my captive audience to do so now. My apologies for inflicting a very theoretical discussion on you.

The theory holds, in the most general terms, that the way philosophy and natural science have taught us to look at the world is basically inappropriate when we are considering events which have thinking participants. Both philosophy and natural science have gone to great lengths to separate events from the observations which relate to them. Events are facts and observations are true or false, depending on whether or not they correspond to the facts.

This way of looking at things can be very productive. The achievements of natural science are truly awesome, and the separation between fact and statement provides a very reliable criterion of truth. So I am in no way critical of this approach. The separation between fact and statement was probably a greater advance in the field of thinking than

the invention of the wheel in the field of transportation.

But exactly because the approach has been so successful, it has been carried too far. Applied to events which have thinking participants, it provides a distorted picture of reality. The key feature of these events is that the participants' thinking affects the situation to which it refers. Facts and thoughts cannot be separated in the same way as they are in natural science or, more exactly, by separating them we introduce a distortion which is not present in natural science, because in natural science thoughts and statements are outside the subject matter, whereas in the social sciences they constitute part of the subject matter. If the study of events is confined to the study of facts, an important element, namely, the participants' thinking, is left out of account. Strange as it may seem, that is exactly what has happened, particularly in economics, which is the most scientific of the social sciences.

Classical economics was modeled on Newtonian physics. It sought to establish the equilibrium position and it used differential equations to do so. To make this intellectual feat possible, economic theory assumed perfect knowledge on the part of the participants. Perfect knowledge meant that the participants' thinking corresponded to the facts and therefore it could be ignored. Unfortunately, reality never quite conformed to the theory. Up to a point, the discrepancies could be dismissed by saying that the equilibrium situation represented the final outcome and the divergence from equilibrium represented temporary noise. But, eventually, the assumption of perfect knowledge became untenable and it was replaced by a methodological device which was invented by my professor at the London School of Economics, Lionel Robbins, who asserted that the task of economics is to study the relationship between supply and demand; therefore it must take supply and demand as given. This methodological device has managed to protect equilibrium theory from the onslaught of reality down to the present day.

I don't know too much about the prevailing theory about financial markets but, from what little I know, it continues to maintain the approach established by classical economics. This means that financial markets are envisaged as playing an essentially passive role; they discount the future and they do so with remarkable accuracy. There is some kind of magic involved and that is, of course, the magic of the marketplace where all the participants, taken together, are endowed with an intelligence far superior to that which could be attained by any particular individual. I think this interpretation of the way financial markets operate is severely distorted. That is why I have not bothered to familiarize myself with efficient market theory and modern portfolio theory, and that is why I take such a jaundiced view of derivative instruments which are based on what I consider a fundamentally flawed principle. Another reason is that I am rather poor in mathematics.

It may seem strange that a patently false theory should gain such widespread acceptance, except for one consideration; that is, that all our theories about social events are distorted in some way or another. And that is the starting point of my theory, the theory of reflexivity, which holds that our thinking is inherently biased. Thinking participants cannot act on the basis of knowledge. Knowledge presupposes facts which occur

independently of the statements which refer to them; but being a participant implies that one's decisions influence the outcome. Therefore, the situation participants have to deal with does not consist of facts independently given but facts which will be shaped by the decision of the participants. There is an active relationship between thinking and reality, as well as the passive one which is the only one recognized by natural science and, by way of a false analogy, also by economic theory.

I call the passive relationship the "cognitive function" and the active relationship the "participating function," and the interaction between the two functions I call "reflexivity." Reflexivity is, in effect, a two-way feedback mechanism in which reality helps shape the participants' thinking and the participants' thinking helps shape reality in an unending process in which thinking and reality may come to approach each other but can never become identical. Knowledge implies a correspondence between statements and facts, thoughts and reality, which is not possible in this situation. The key element is the lack of correspondence, the inherent divergence, between the participants' views and the actual state of affairs. It is this divergence, which I have called the "participant's bias," which provides the clue to understanding the course of events. That, in very general terms, is the gist of my theory of reflexivity.

The theory has far-reaching implications. It draws a sharp distinction between natural science and social science, and it introduces an element of indeterminacy into social events which is missing in the events studied by natural science. It interprets social events as a never-ending historical process and not as an equilibrium situation. The process cannot be explained and predicted with the help of universally valid laws, in the manner of natural science, because of the element of indeterminacy introduced by the participants' bias. The implications are so far-reaching that I can't even begin to enumerate them. They range from the inherent instability of financial markets to the concept of an open society which is based on the recognition that nobody has access to the ultimate truth. The theory gives rise to a new morality as well as a new epistemology. As you probably know, I am the founder—and the funder—of the Open Society Foundation. That is why I feel justified in claiming that the theory of reflexivity has guided me both in making and in spending money.

But is it possible to come up with a valid new theory about the relationship between thinking and reality? It seems highly unlikely. The subject has been so thoroughly explored that probably everything that can be said has been said. In my defense, I did not produce the theory in a vacuum. The logical indeterminacy of self-referring statements was first discussed by Epimenides, the Cretan philosopher, who said, "Cretans always lie," and the paradox of the liar was the basis of Bertrand Russell's theory of classes. But I am claiming more than a logical indeterminacy. Reflexivity is a two-way feedback mechanism, which is responsible for a causal indeterminacy as well as a logical one. The causal indeterminacy resembles Heisenberg's uncertainty principle, but there is a major difference: Heisenberg's theory deals with observations, whereas reflexivity deals with the role of thinking in generating observable phenomena.

I am thrilled by the possibility that I may have reached a profound new insight, but I am

also scared because such claims are usually made by insane people and there are many more insane people in the world than there are people who have reached a profound new insight. I wonder whether my insight has an objective validity or only a subjective significance.

That is why I am so eager to submit my ideas to a critical examination and that is why I find the present situation, where I am taken so seriously but my theory is not, so frustrating. As I have said before, the theory of reflexivity has received practically no serious consideration. It is treated as the self-indulgence of a man who made a lot of money in the stock market. It is generally summed up by saying that markets are influenced by psychological factors, and that is pretty trite. But that is not what the theory says. It says that, in certain cases, the participants' bias can change the fundamentals which are supposed to determine market prices.

I ask myself, why did I fail to communicate this point? The answer I come up with is that I tried to say too much, too soon. I tried to propound a general theory of reflexivity at a time when reflexivity as a phenomenon is not even recognized. In retrospect, I think I should have started more modestly; I should have tried to prove the existence of reflexivity as a phenomenon before I tried to revise our view of the world based on that phenomenon. It can be done relatively easily, and the financial markets provide an excellent laboratory in which to do it. And that is what I should like to do here today.

What I need to do is to demonstrate that there are instances where the participants' bias is capable of affecting not only market prices but also the so-called fundamentals that market prices are supposed to reflect. I have collected and analyzed such instances in The Alchemy of Finance, so all I need to do here is simply to enumerate them. In the case of stocks, I have analyzed two particular instances which demonstrate my case perfectly; one is the conglomerate boom and bust of the late 1960s, and the other is the boom and bust of real estate investment trusts in the early 70s. I cite may other instances, such as the leveraged buyout boom of the 1980s and the boom/bust sequences engendered by foreign investors. But these cases are less clear cut.

The common thread in the two instances I have mentioned is so-called equity leveraging; that is to say, companies can use inflated expectations to issue new stock at inflated prices, and the resulting increase in earnings per share can go a long way to validate the inflated expectations. But equity leveraging is only one of many possible mechanisms for transmitting the participants' bias to the underlying fundamentals. Consider, for instance, the boom in international lending which occurred in the 1970s and led to the bust of 1982. In the boom, banks relied on so-called debt ratios, which they considered as objective measurements of the ability of the borrowing countries to service their debt, and it turned out that these debt ratios were themselves influenced by the lending activity of the banks.

In all these cases, the participants' bias involved an actual fallacy: in the case of the conglomerate and mortgage trust booms, the growth in earnings per share was treated as if it were independent of equity leveraging; and in the case of the international lending

boom, the debt ratio was treated as if it were independent of the lending activities of the banks. But there are other cases where no such fallacy is involved. For instance, in a freely-fluctuating currency market, a change in exchange rates has the capacity to affect the so-called fundamentals which are supposed to determine exchange rates, such as the rate of inflation in the countries concerned; so that any divergence from a theoretical equilibrium has the capacity to validate itself. This self-validating capacity encourages trend-following speculation, and trend-following speculation generates divergences from whatever may be considered the theoretical equilibrium. The circular reasoning is complete. The outcome is that freely-fluctuating currency markets tend to produce excessive fluctuations and trend-following speculation tends to be justified.

I believe that these examples are sufficient to demonstrate that reflexivity is real; it is not merely a different way of looking at events; it is a different way in which events unfold. It doesn't occur in every case but, when it does, it changes the character of the situation. Instead of a tendency towards some kind of theoretical equilibrium, the participants' views and the actual state of affairs enter into a process of dynamic disequilibrium which may be mutually self-reinforcing at first, moving both thinking and reality in a certain direction. but is bound to become unsustainable in the long run and engender a move in the opposite direction. The net result is that neither the participants' views nor the actual state of affairs returns to the condition from which it started. Once the phenomenon of reflexivity has been isolated and recognized, it can be seen to be at work in a wide variety of situations. I studied one such situation in The Alchemy of Finance which was particularly relevant at the time the book was written. I called it "Reagan's Imperial Circle." It consisted of financing a massive armaments program with money borrowed from abroad, particularly from Japan. I showed that the process was initially selfreinforcing but it was bound to become unsustainable. A similar situation has arisen recently with the reunification of Germany, which eventually led to the breakdown of the European Exchange Rate Mechanism. The ERM operated in near- equilibrium conditions for about a decade before the reunification of Germany created a dynamic disequilibrium.

What renders reflexivity significant is that it occurs only intermittently. If it were present in all situations all the time, it would merely constitute a different way of looking at events and not a different way for events to evolve. That is the point I failed to make sufficiently clear in my book. I presented my theory of reflexivity as a general theory in which the absence of reflexivity appears as a special case. I was, of course, trying to imitate Keynes, who proposed his general theory of employment in which full employment was a special case. But Keynes proposed his theory when unemployment was a well-established fact, whereas I proposed the theory of reflexivity before the phenomenon has been recognized. In doing so, I both overstated and understated my case. I overstated it by arguing that the methods and criteria of the natural sciences are totally inapplicable to the study of social phenomena. I called social science a false metaphor. That is an exaggeration because there are many normal, everyday, repetitive situations which can be explained and predicted by universally valid laws whose validity can be tested by scientific method. And even historical, reflexive processes have certain repetitive aspects which lend themselves to statistical generalizations. For instance, the trade cycle follows a certain repetitive pattern, although each instance may have some unique features and there is a lot more to be gained from understanding the unique features than the repetitive pattern.

I have also understated my case by presenting reflexivity as a different way of looking at the structure of social events rather than a different way in which events unfold when reflexivity comes into play. I made the point that, in natural science, one set of facts follows another irrespective of what anybody thinks; whereas in the events studied by social science, there is a two-way interaction between perception and facts. I also drew a distinction between humdrum, everyday events in which the element of indeterminacy introduced by the reflexive connection can be treated as mere noise, and historical events where the reflexive interaction brings about an irreversible change both in the participants' views and the actual state of affairs. All this is very profound and very significant, but the really interesting undertaking is to study the difference between humdrum and historical events and to gain a better understanding of historical processes.

I have done a lot of work in that direction since I wrote The Alchemy of Finance, not so much in the financial markets as in the historical arena. I have come to distinguish between normal conditions and far-from-equilibrium conditions. In normal conditions, there is a tendency for the participants' views and the actual state of affairs to converge or, at least, there are mechanisms at work to prevent them from drifting too far apart. I call these conditions "normal," because that is what our intellectual traditions—including philosophy and scientific method —have prepared us for. I contrast them with far-from-equilibrium conditions, where the participants' views are far removed from the actual state of affairs and there is no tendency for the two of them to come together. I have always found the far-from-equilibrium conditions much more fascinating, and I have studied them both in theory and in practice.

There are two very different kinds of far-from-equilibrium conditions: one is associated with the absence of change, and the other with revolutionary change. These two opposite poles act as "strange attractors"—an expression with which has become familiar since chaos theory has come into vogue.

So we can observe three very different conditions in history: the "normal," in which the participants' views and the actual state of affairs tend to converge; and two far-from-equilibrium conditions, one of apparent changelessness, in which thinking and reality are very far apart and show no tendency to converge, and one of revolutionary change in which the actual situation is so novel and unexpected and changing so rapidly that the participants' views cannot keep up with it.

Interestingly, the rise and fall of the Soviet system presents both extremes. During Stalin's time, reality and dogma were very far apart, but both of them were very rigid and showed no tendency to come together. Indeed, the divergence increased with the passage of time. When the system finally collapsed, people could not cope with the pace of change and events spun out of control. That is what we have witnessed recently.

But the two extremes can also be observed in totally unrelated contexts. Take, for instance, the banking industry in the United States. After the breakdown of the banking system in the Great Depression, it became closely regulated and very rigid; but when the restrictions were relaxed, the industry swung to the other extreme and entered a period of revolutionary change. I can locate the transition point with great precision: it was on that evening in 1973 when the management of First National City Bank held an unprecedented meeting for securities analysts in order to promote the stock as a growth stock. The pattern in the rise and fall of the Soviet system closely parallels the pattern in the fall and rise of the American banking system.

These three conditions are perhaps better explained by using an analogy. The analogy is with water, which also can be found in nature in three conditions: as a liquid, a solid or a gas. The three historical conditions I am trying to describe are as far apart as water, ice and steam. In the case of H2O, we can define exactly the three conditions; it has to do with temperature. Can we establish a similar demarcation line among the three conditions of historical change? I believe we can, and it has to do with the values that guide people in their actions. But I am not yet ready to give a firm answer. That is the problem that I am currently working on. But I feel rather exposed in dealing with such an esoteric issue. I need to know whether what I have said so far makes any sense; that is why I have imposed on you by giving you this rather heavy theoretical lecture, and I would welcome your comments either here or on another occasion.