FEEDBACK

It is the fundamental principle that underlies all self-regulating systems, not only machines but also the processes of life and the tides of human affairs

by Arnold Tustin

FOR hundreds of years a few examples of true automatic control systems have been known. A very early one was the arrangement on windmills of a device to keep their sails always facing into the wind. It consisted simply of a miniature windmill which could rotate the whole mill to face in any direction. The small mill's sails were at right angles to the main ones, and whenever the latter faced in the wrong direction, the wind caught the small sails and rotated the mill to the correct position. With steam power came other automatic mechanisms: the engine-governor, and then the steering servo-engine on ships, which operated the rudder in correspondence with movements of the helm. These devices, and a few others such as simple voltage regulators, constituted man's achievement in automatic control up to about 20 years ago.

In the past two decades necessity, in the form of increasingly acute problems arising in our ever more complex technology, has given birth to new families of such devices. Chemical plants needed regulators of temperature and flow; air warfare called for rapid and precise control of searchlights and anti-aircraft guns; radio required circuits which would give accurate amplification of signals.

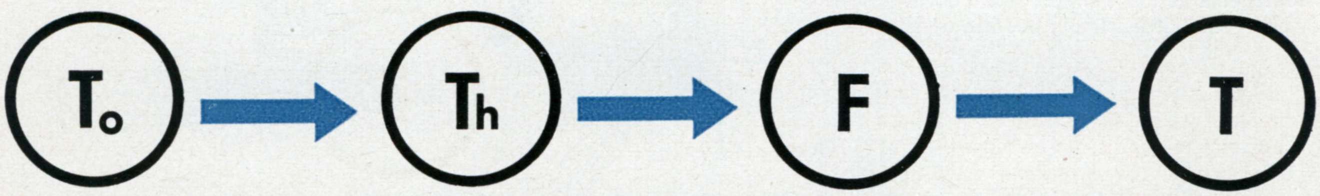
Thus the modern science of automatic control has been fed by streams from many sources. At first, it now seems surprising to recall, no connection between these various developments was recognized. Yet all control and regulating systems depend on common principles. As soon as this was realized, progress became much more rapid. Today the design of controls for a modern boiler or a guided missile, for example, is based largely on principles first developed in the design of radio amplifiers.

Indeed, studies of the behavior of automatic control systems give us new insight into a wide variety of happenings in nature and in human affairs. The notions that engineers have evolved from these studies are useful aids in understanding how a man stands upright without toppling over, how the human heart beats, why our economic system suffers from slumps and booms, why the rabbit population in parts of Canada regularly fluctuates between scarcity and abundance.

The chief purpose of this article is to make clear the common pattern that underlies all these and many other varied phenomena. This common pattern is the existence of feedback, or—to express the same thing rather more generally—interdependence.

We should not be able to live at all, still less to design complex control systems, if we did not recognize that there are regularities in the relationship between events—what we call "cause and effect." When the room is warmer, the thermometer on the wall reads higher. We do not expect to make the room warmer by pushing up the mercury in the thermometer. But now consider the case when the instrument on the wall is not a simple thermometer but a thermostat, contrived so that as its reading goes above a chosen setting, the fuel supply to the furnace is progressively reduced, and, conversely, as its reading falls below that setting, the fuel flow is increased. This is an example of a familiar control system. Not only does the reading of the thermometer depend on the warmth of the room, but the warmth of the room also depends on the reading of the thermometer. The two quantities are interdependent. Each is a cause, and each an effect, of the other. In such cases we have a closed chain or sequence—what engineers call a "closed loop" (*see diagram on the opposite page*).

In analyzing engineering and scientific problems it is very illuminating to sketch out first the scheme of dependence and see how the various quantities involved in the problem are determined by one another and by disturbances from outside the system. Such a diagram enables one to tell at a glance whether a system is an open or a closed one. This is an important distinction, because a closed system possesses several significant properties. Not only can it act as a regulator, but it is capable of



OPEN SEQUENCE of control is illustrated by a system for regulating the temperature of a room. T_o is a variation in the temperature outdoors. Th is the variation of a thermometer. F is the fuel control of a furnace. T is the variation of the temperature in the room. In such a system of control there is no feedback.

various "self-excitatory" types of behavior—like a kitten chasing its own tail.

The now-popular name for this process is "feedback." In the case of the thermostat, the thermometer's information about the room temperature is fed back to open or close the valve, which in turn controls the temperature. Not all automatic control systems are of the closed-loop type. For example, one might put the thermometer outside in the open air, and connect it to work the fuel valve through a specially shaped cam, so that the outside temperature regulates the fuel flow. In this open-sequence system the room temperature has no effect; there is no feedback. The control compensates only that disturbance of room temperature caused by variation of the outdoor temperature. Such a system is not necessarily a bad or useless system; it might work very well under some circumstances. But it has two obvious shortcomings. Firstly, it is a "calibrated" system; that is to say, its correct working would require careful preliminary testing and special shaping of the cam to suit each particular application. Secondly, it could not deal with any but standard conditions. A day that was windy as well as cold would not get more fuel on that account.

The feedback type of control avoids these shortcomings. It goes directly to the quantity to be controlled, and it corrects indiscriminately for all kinds of disturbance. Nor does it require calibration for each special condition.

Feedback control, unlike open-sequence control, can never work without *some* error, for the error is depended upon to bring about the correction. The objective is to make the error as small as possible. This is subject to certain limitations, which we must now consider.

The principle of control by feedback is quite general. The quantities that it may control are of the most varied kinds, ranging from the frequency of a national electric-power grid to the degree of anesthesia of a patient under surgical operation. Control is exercised by negative feedback, which is to say that the information fed back is the amount of departure from the desired condition.

ANY QUANTITY may be subjected to control if three conditions are met. First, the required changes must be controllable by some physical means, a regulating organ. Second, the controlled quantity must be measurable, or at least comparable with some standard; in other words, there must be a measuring device. Third, both regulation and measurement must be rapid enough for the job in hand.

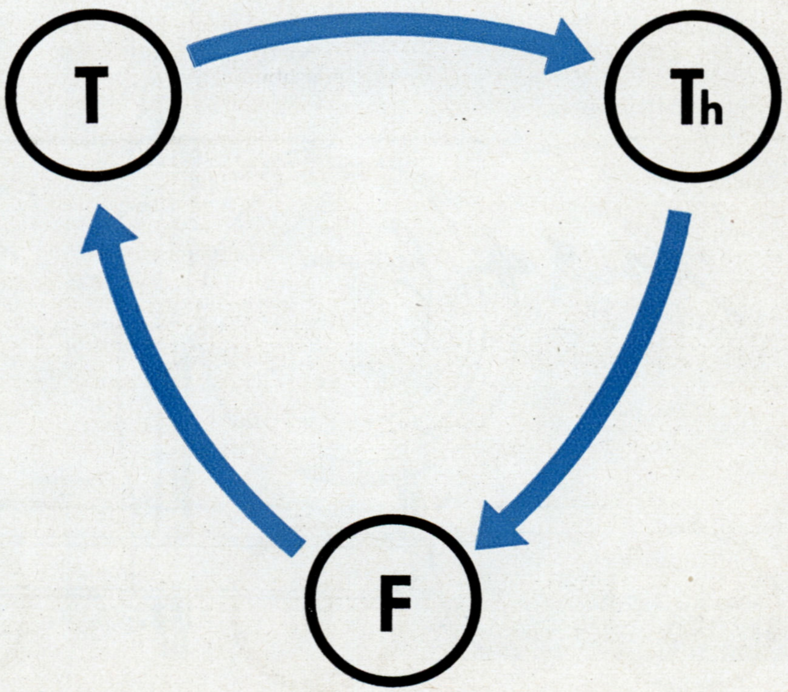
As an example, take one of the simplest and commonest of industrial requirements: to control the rate of flow of liquid along a pipe. As the regulating organ we can use a throttle valve, and as the measuring device, some form of flowmeter. A signal from the flowmeter, telling the actual rate of flow through the pipe, goes to the "controller"; there it is compared with a setting giving the required rate of flow. The amount and direction of "error," *i.e.*, deviation from this setting, is then transmitted to the throttle valve as an operating signal to bring about adjustment in the required direction (*see diagram at the top of page 53*).

In flow-control systems the signals are usually in the form of variations in air pressure, by which the flowmeter measures the rate of flow of the liquid. The pressure is transmitted through a small-bore pipe to the controller, which is essentially a balance piston. The difference between this received pressure and the setting regulates the air pressure in another pipeline that goes to the regulating valve.

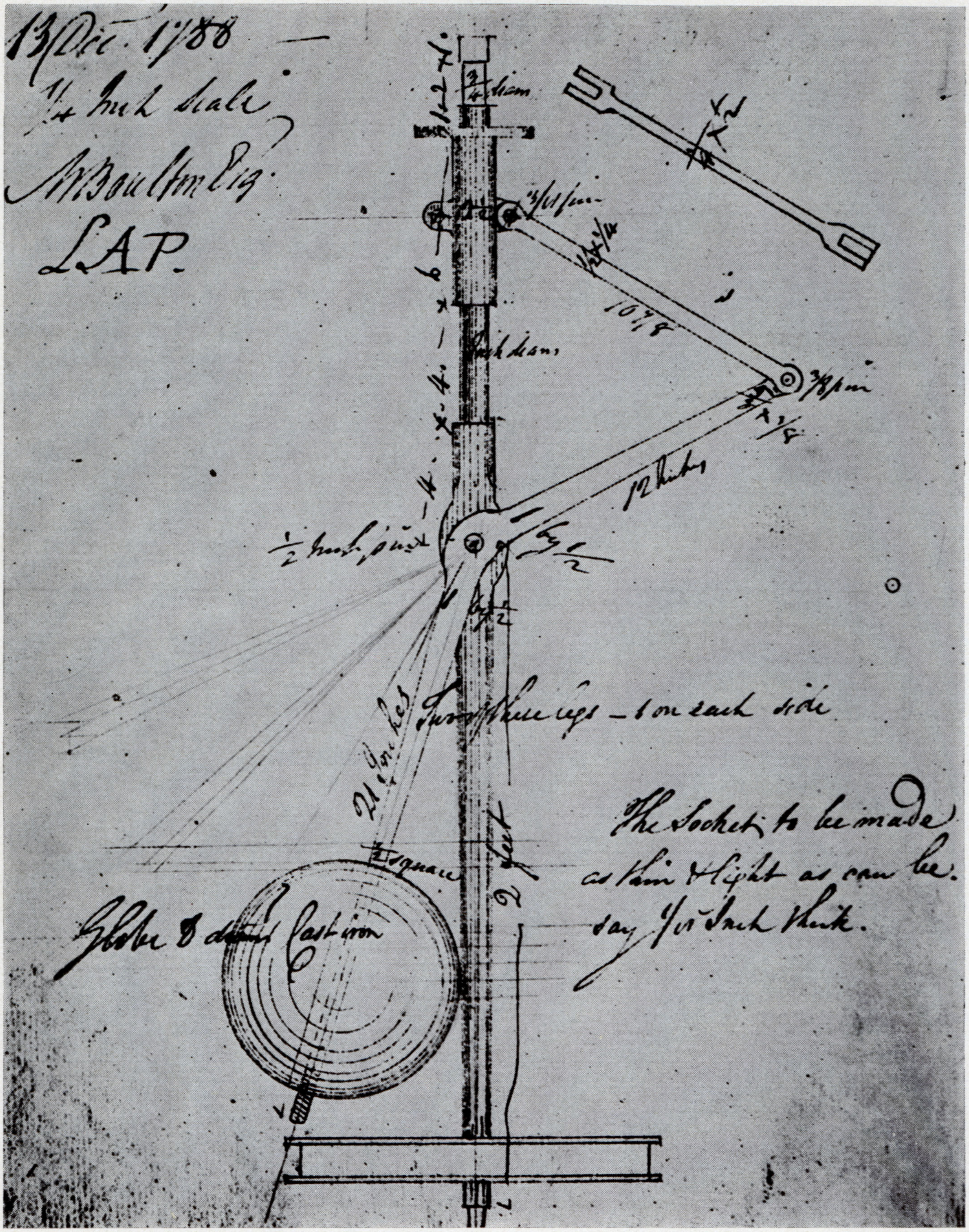
Signals of this kind are slow, and difficulties arise as the system becomes complex. When many controls are concentrated at a central point, as is often the case, the air-pipes that transmit the signals may have to be hundreds of feet long, and pressure changes at one end reach the other only after delays of some seconds. Meanwhile the error may have become large. The time-delay often creates another problem: overcorrection of the error, which causes the system to oscillate about the required value instead of settling down.

For further light on the principles involved in control systems let us consider the example of the automatic gun-director. In this problem a massive gun must be turned with great precision to angles indicated by a fly-power pointer on a clock-dial some hundreds of feet away. When the pointer moves, the gun must turn correspondingly. The quantity to be controlled is the angle of the gun. The reference quantity is the angle of the clock-dial pointer. What is needed is a feedback loop which constantly compares the gun angle with the pointer angle and arranges matters so that if the gun angle is too small, the gun is driven forward, and if it is too large, the gun is driven back.

The key element in this case is some device which will detect the error of angular alignment between two shafts remote from each other, and which does not require more force than is available at the fly-power transmitter shaft. There are several kinds of electrical elements that will serve such a purpose. The one usually selected is a pair of the miniature alternating-current machines known as selsyns. The two selsyns, connected respectively to the transmitter shaft and the gun, provide an electrical signal proportional to the error of alignment. The signal is amplified and fed to a generator which in turn feeds a motor that



CLOSED SEQUENCE of control is illustrated by a system for regulating the temperature of a room by means of a thermostat. Here Th is a thermostat rather than a thermometer. In such a system there is feedback.



EARLIEST KNOWN DRAWING of the flyball-governor was made in 1788. The governor was invented by James Watt. At the upper left appear the date and the name of Watt's associate Matthew Boulton. Only half of a governor is shown in the drawing, but below the center are the words: "Two of these legs—1 on each side." Later Watt attempted to prevent the oscillation of the governor by fitting it with stops for the balls, one to keep them from coming too close together and the other to prevent them from "opening too wide asunder."

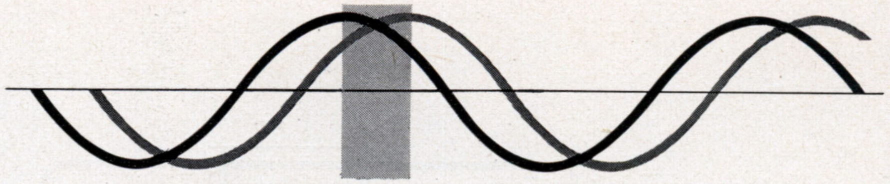
drives the gun (*see diagram on the next page*).

THIS GIVES the main lines of a practicable scheme, but if a system were built as just described, it would fail. The gun's inertia would carry it past the position of correct alignment; the new error would then cause the controller to swing it back, and the gun would hunt back and forth without ever settling down.

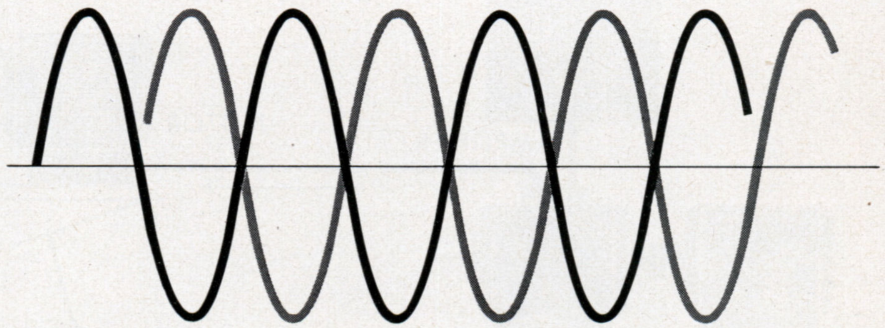
This oscillatory behavior, maintained by "self-excitation," is one of the principal limitations of feedback control. It is the chief enemy of the control-system designer, and the key to progress has been the finding of various simple means to prevent oscillation. Since oscillation is a very general phenomenon, it is worth while to look at the mechanism in detail, for what we learn about oscillation in man-made control systems may suggest means of inhibiting oscillations of other kinds—such as economic booms and slumps, or periodic swarms of locusts.

Consider any case in which a quantity that we shall call the output depends on another quantity we shall call the input. If the input quantity oscillates in value, then the output quantity also will oscillate, not simultaneously or necessarily in the same way, but with the same frequency. Usually in physical systems the output oscillation lags behind the input. For example, if one is boiling water and turns the gas slowly up and down, the amount of steam increases and decreases the same number of times per minute, but the maximum amount of steam in each cycle must come rather later than the maximum application of heat, because of the time required for heating. If the first output quantity in turn affects some further quantity, the variation of this second quantity in the sequence will usually lag still more, and so on. The lag (as a proportion of one oscillation) also usually increases with frequency—the faster the input is varied, the farther behind the output falls.

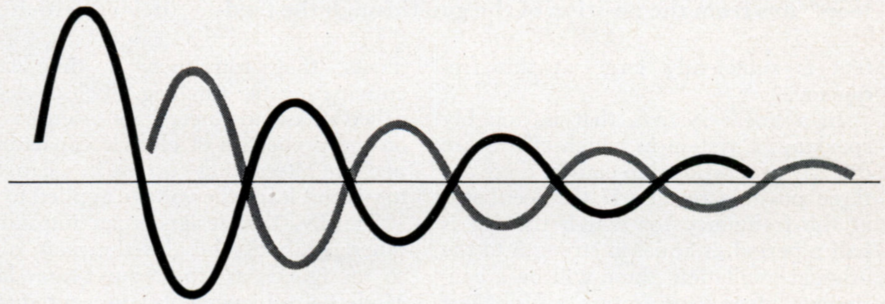
Now suppose that in a feedback system some quantity in the closed loop is oscillating. This causes the successive quantities around the loop to oscillate also. But the loop comes around to the original quantity, and we have here the mechanism by which an oscillation may maintain itself. To see how this can happen, we must remember that with the feedback negative, the motion it causes would be opposite to the original motion, if it were not for the lags. It is only when the lags add up to just half a cycle that the feedback maintains the assumed motion. Thus any system with negative feedback will maintain a continuous oscillation when disturbed if (a) the time-delays in response at some frequency add up to half a period of oscillation, and (b) the feedback ef-



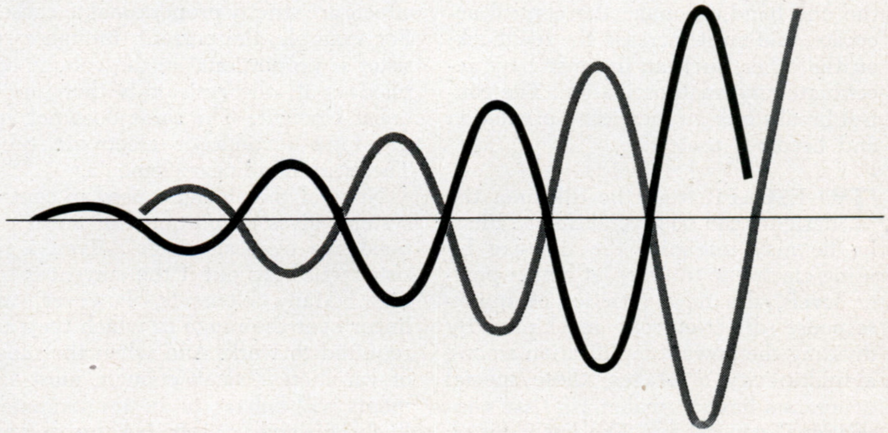
REGULAR OSCILLATORY VARIATION of a quantity put into a feedback system (*black curve*) is followed by a similar variation in the output quantity (*gray curve*). The gray rectangle indicates the time-delay.



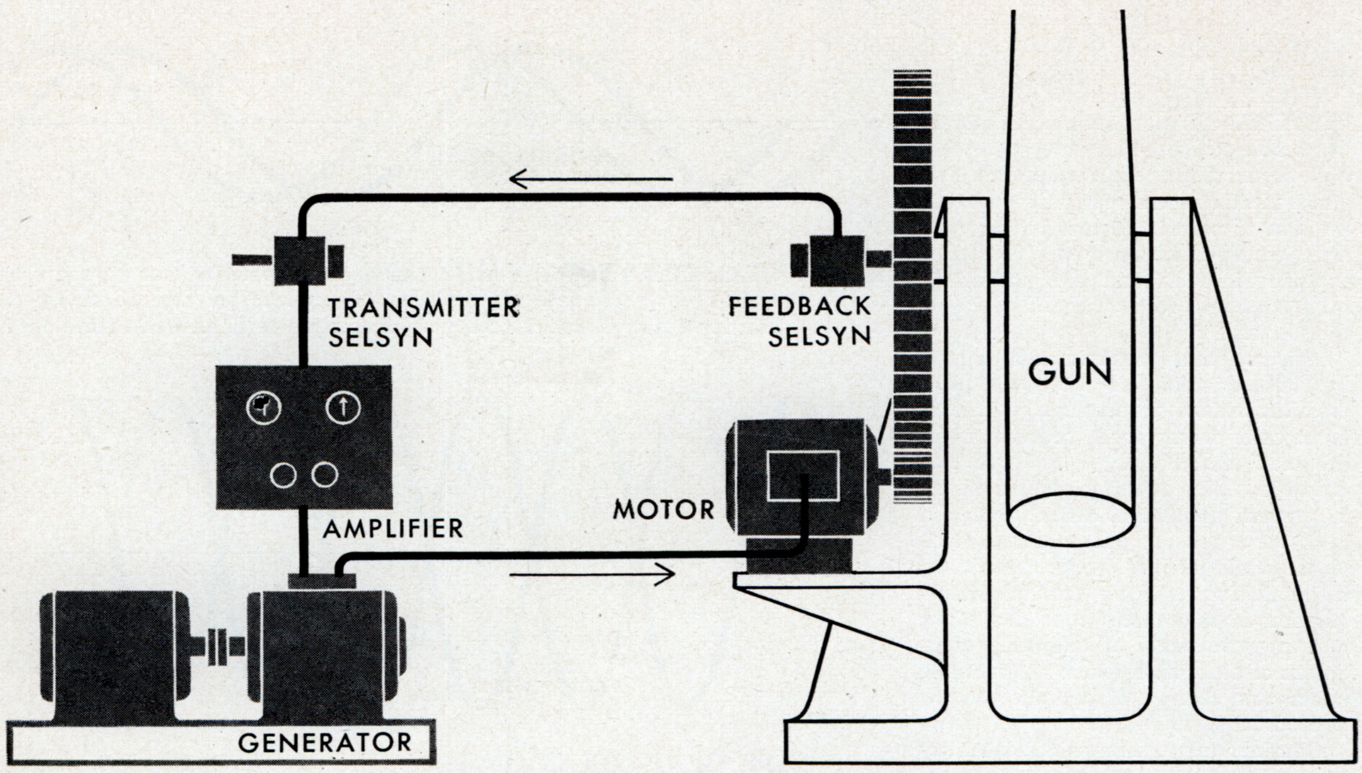
ONE TYPE OF OSCILLATION occurs when the feedback (*gray curve*) of a system is equal and opposed to its error (*black curve*). Here the term error is used to mean any departure of the system from its desired state.



SECOND TYPE OF OSCILLATION occurs when the feedback (*black curve*) of a system is less than and opposed to the error (*gray curve*). This set of conditions tends to damp the disturbance in the system.



THIRD TYPE OF OSCILLATION occurs when the feedback (*gray curve*) of a system is greater than and opposed to the error (*black curve*). This set of conditions tends to amplify the disturbance in the system.



ELEVATION OF A GUN is controlled by an electrical feedback system. The closed sequence, or closed loop, runs from the position of the gun through the feedback selsyn, the transmitter selsyn, the amplifier and the generator to the motor. The selsyn is an electrical device which transmits position or speed of rotation.

fect is sufficiently large at this frequency.

In a linear system, that is, roughly speaking, a system in which effects are directly proportional to causes, there are three possible results. If the feedback, at the frequency for which the lag is half a period, is equal in strength to the original oscillation, there will be a continuous steady oscillation which just sustains itself. If the feedback is greater than the oscillation at that frequency, the oscillation builds up; if it is smaller, the oscillation will die away.

This situation is of critical importance for the designer of control systems. On the one hand, to make the control accurate, one must increase the feedback; on the other, such an increase may accentuate any small oscillation. The control breaks into an increasing oscillation and becomes useless.

TO ESCAPE from the dilemma the designer can do several things. Firstly, he may minimize the time-lag by using electronic tubes or, at higher power levels, the new varieties of quick-response direct-current machines. By dividing the power amplification among a multiplicity of stages, these special generators have a smaller lag than conventional generators. The lag is by no means negligible, however.

Secondly, and this was a major advance in the development of control systems, the designer can use special elements that introduce a time-lead, anticipating the time-lag. Such devices, called phase-advancers, are often based on the properties of electric capacitors, because alternating current in a capacitor circuit leads the voltage applied to it.

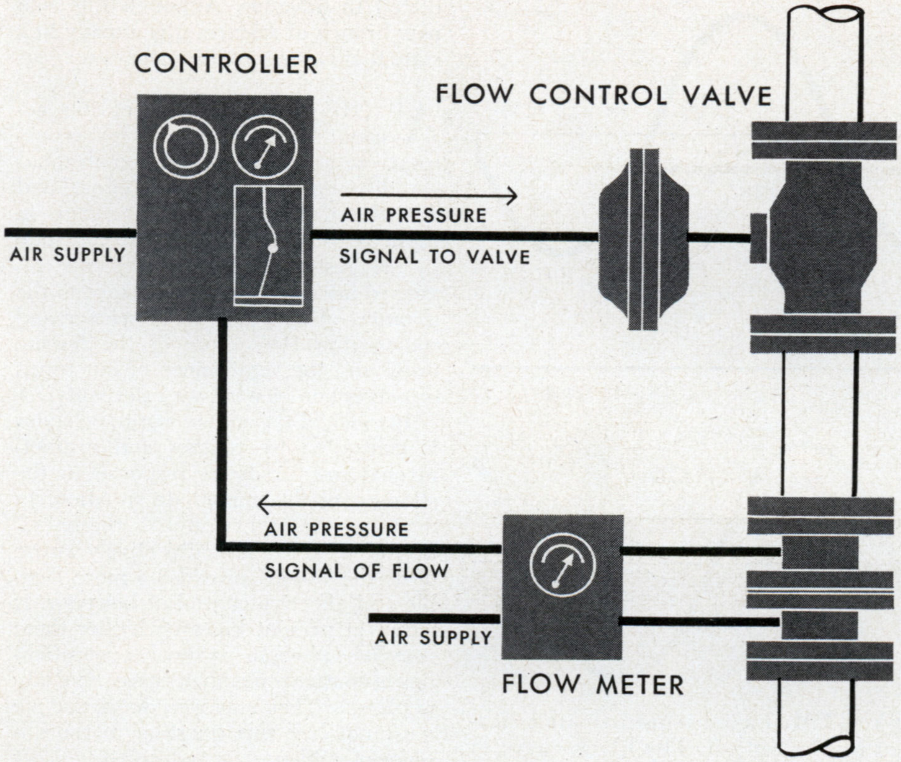
Thirdly, the designer can introduce other feedbacks besides the main one, so designed as to reduce time-lag. Modern achievements in automatic control are based on the use of combinations of such devices to obtain both accuracy and stability.

So far we have been treating these systems as if they were entirely linear. A system is said to be linear when all effects are strictly proportional to causes. For example, the current through a resistor is proportional to the voltage applied to it; the resistor is therefore a linear element. The same does not apply to a rectifier or electronic tube. These are non-linear elements.

None of the elements used in control systems gives proportional or linear dependence over all ranges. Even a resistor will burn out if the current is too high. Many elements, however, are linear over the range in which they are required to work. And when the range of variation is small enough, most elements will behave in an approximately linear fashion, simply because a very small bit of a curved graph does not differ significantly from a straight line.

We have seen that linear closed-sequence systems are delightfully simple to understand and—even more important—very easy to handle in exact mathematical terms. Because of this, most introductory accounts of control systems either brazenly or furtively assume that all such systems are linear. This gives the rather wrong impression that the principles so deduced may have little application to real, non-linear, systems. In practice, however, most of the characteristic behavior of control systems is affected only in detail by the non-linear nature of the dependences. It is essential to be clear that non-linear systems are not excluded from feedback control. Unless the departures from linearity are large or of special kinds, most of what has been said applies with minor changes to non-linear systems.

LONG BEFORE man existed, evolution hit upon the need for anti-oscillating features in feedback control and incorporated them in the body mechanisms of the animal world. Signals in the animal body are transmitted by trains of pulses along nerve fibers. When a sensory organ is stimulated, the stimulus will produce pulses at a greater rate if it is increasing than if it is decreasing. The pattern of nerve response to an oscillatory stimulus is shown in the diagram on page 54. The maximum response, or output signal, occurs before the maximum of the stimulus. This is just the anticipatory type of effect (the time-lead) that is required for high-



RATE OF FLOW IN A PIPE is controlled by a pneumatic feedback system. Here the closed loop runs from the flow of fluid in the pipe through the flow meter and the recording controller to the flow-control valve.

accuracy control. Physiologists now believe that the anticipatory response has evolved in the nervous system for, at least in part, the same reason that man wants it in his control mechanisms—to avoid overshooting and oscillation. Precisely what feature of the structure of the nerve mechanism gives this remarkable property is not yet fully understood.

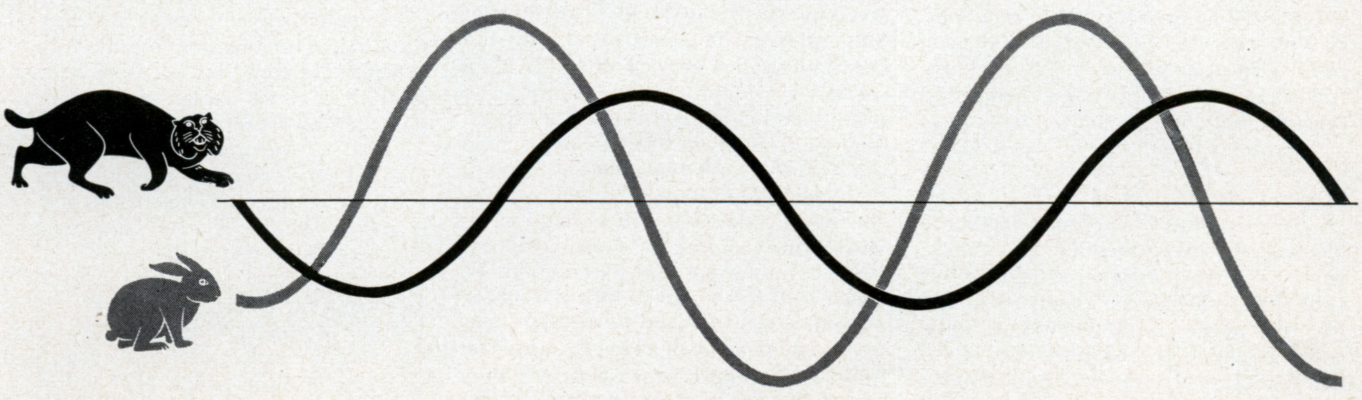
Fascinating examples of the consequences of interdependence arise in the fluctuations of animal populations in a given territory. These interactions are sometimes extremely complicated. Charles Darwin invoked such a scheme to explain why there are more bumblebees near towns. His explanation was that near towns there are more cats; this means fewer field mice, and field mice are the chief ravagers of bees' nests. Hence near towns bees enjoy more safety.

The interdependence of animal species sometimes produces a periodic oscillation. Just to show how this can happen, and leaving out complications that are always present in an actual situation, consider a territory inhabited by rabbits and lynxes, the rabbits being the chief food of the lynxes. When rabbits are abundant, the lynx population will increase. But as the lynxes become abundant, the rabbit population falls, because more rabbits are caught. Then as the rabbits diminish, the lynxes go hungry and decline. The result is a self-maintaining oscillation, sustained by negative feedback with a time-delay (*see diagram below*).

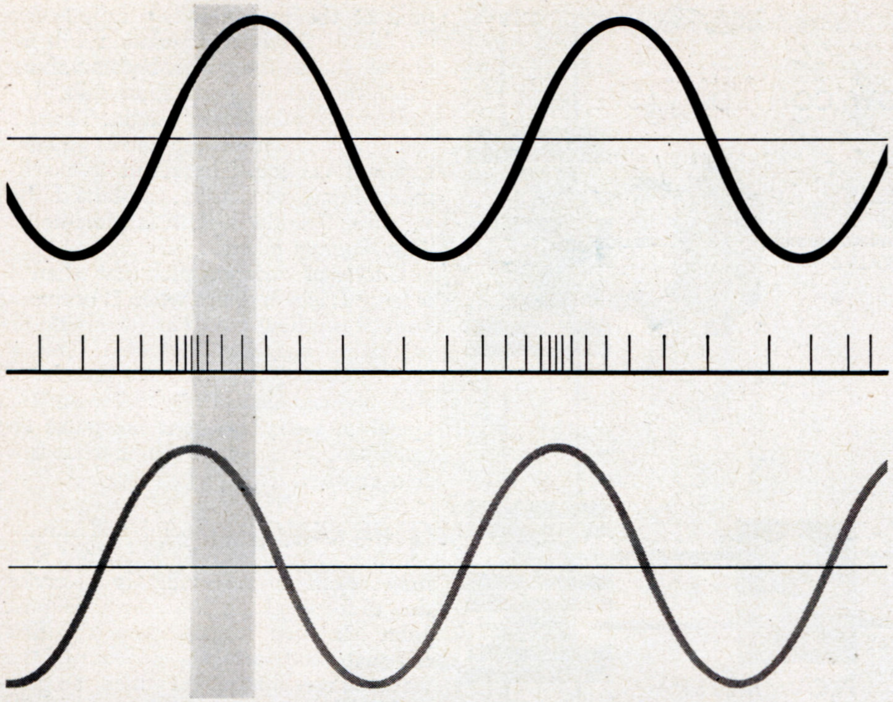
Curves of variation such that when R is large L is rising, but when L becomes large R is falling must have the periodic oscillatory character indicated. This is not, of course, the complete picture of such phenomena as the well-known "fur cycle" of Canada, but it illustrates an important element in the mechanisms that cause it.

THE PERIODIC booms and slumps in economic activity stand out as a major example of oscillatory behavior due to feedback. In 1935 the economist John Maynard Keynes gave the first adequate and satisfying account of the essential mechanisms on which the general level of economic activity depends. Although Keynes did not use the terminology of control-system theory, his account fits precisely the same now-familiar pattern.

Keynes' starting point was the simple notion that the level of economic activity depends on the rate at which goods are bought. He took the essential further step of distinguishing two kinds of buying—of consumption goods and of capital goods. The latter is the same thing as the rate of investment. The money available to buy all these goods is not automatically provided by the wages and profits disbursed in making them, because normally some of this money is saved. The system would therefore run down and stop if it were not for the constant injection of extra demand in the form of new investment. Therefore the level of economic activity and employment depends on the rate of invest-



RABBIT AND LYNX population cycles are an example of a feedback system in nature. Here a fall in the relatively small population of lynxes (*black curve*) is followed by a rise in the large population of rabbits (*gray curve*). This is followed by a rise in the lynx population, a fall in the rabbit population and so on.



FEEDBACK IN THE NERVOUS SYSTEM has a sophisticated feature: the anticipation of the input signal by the output. The black curve at the top of this diagram represents the input signal. The row of vertical lines in the middle indicates the number of nerve pulses in a given time. The gray curve at the bottom translates these pulses into an output signal. The gray rectangle indicates how much the output signal leads the input.

ment. This is the first dependence. The rate of investment itself, however, depends on the expectation of profit, and this in turn depends on the trend, present and expected, of economic activity. Thus not only does economic activity depend on the rate of investment, but the rate of investment depends on economic activity.

Modern theories of the business cycle aim to explain in detail the nature of these dependences and their characteristic non-linearities. This clarification of the mechanisms at work immediately suggests many ways in which, by proper timing of investment expenditure, by more rational business forecasting, and so on, a stable level of optimal economic activity may be achieved in the near future. The day when it can unequivocally be said that slumps belong to the past will certainly be the beginning of a brighter chapter in human history.

THE EXAMPLES of feedback given here are merely a few selected to illustrate general principles. Many more will be described in other articles in this issue. In this article on "theory" I should like to touch on a further point: some ways in which the properties of automatic control systems or other complex feedback systems may be investigated in detail, and their performance perfected.

Purely mathematical methods are remarkably powerful when the system happens to be linear. Sets of linear differential equations are the happy hunting ground of mathematicians. They can turn the equations into a variety of equivalent forms, and generally play tunes on them. For the more general class of non-linear systems, the situation is quite different. There exact determination of the types of motion implied by a set of dependences is usually very laborious or practically impossible.

To determine the behavior of such complex systems two principal kinds of machines are being used. The first is the "analogue" computer. The forms of this type of computer are varied, but they all share a common principle: some system of physical elements is set up with relationships analogous to those existing in the system to be investigated, and the interdependence among them is then worked out in proportional terms. The second kind of aid is the new high-speed digital computer. In this type of machine the quantities are represented by numbers rather than by physical equivalents. The implications of the equations involved are explored by means of arithmetical operations on these numbers. The great speed of operation of these modern machines makes possible calculations that could not be attempted by human computers because of the time required.

The theory of control systems is now so well understood that, with such modern aids, the behavior of even extremely complex systems can be largely predicted in advance. Although this is a new branch of science, it is already in a state that ensures rapid further progress.

AT THE commencement of this account of control systems it was necessary to assume that the human mind can distinguish "cause" and "effect" and describe the regularities of nature in these terms. It may be fitting to conclude by suggesting that the concepts reviewed are not without relevance to the grandest of all problems of science and philosophy: the nature of the human mind and the significance of our forms of perception of what we call reality.

In much of the animal world, behavior is controlled by reflexes and instinct-mechanisms in direct response to the stimulus of the immediate situation. In man and the higher animals the operation of what we are subjectively aware of as the "mind" provides a more flexible and effective control of behavior. It is not at present known whether these conscious phenomena involve potentialities of matter other than those we study in physics. They may well do so, and we must not beg this question in the absence of evidence.

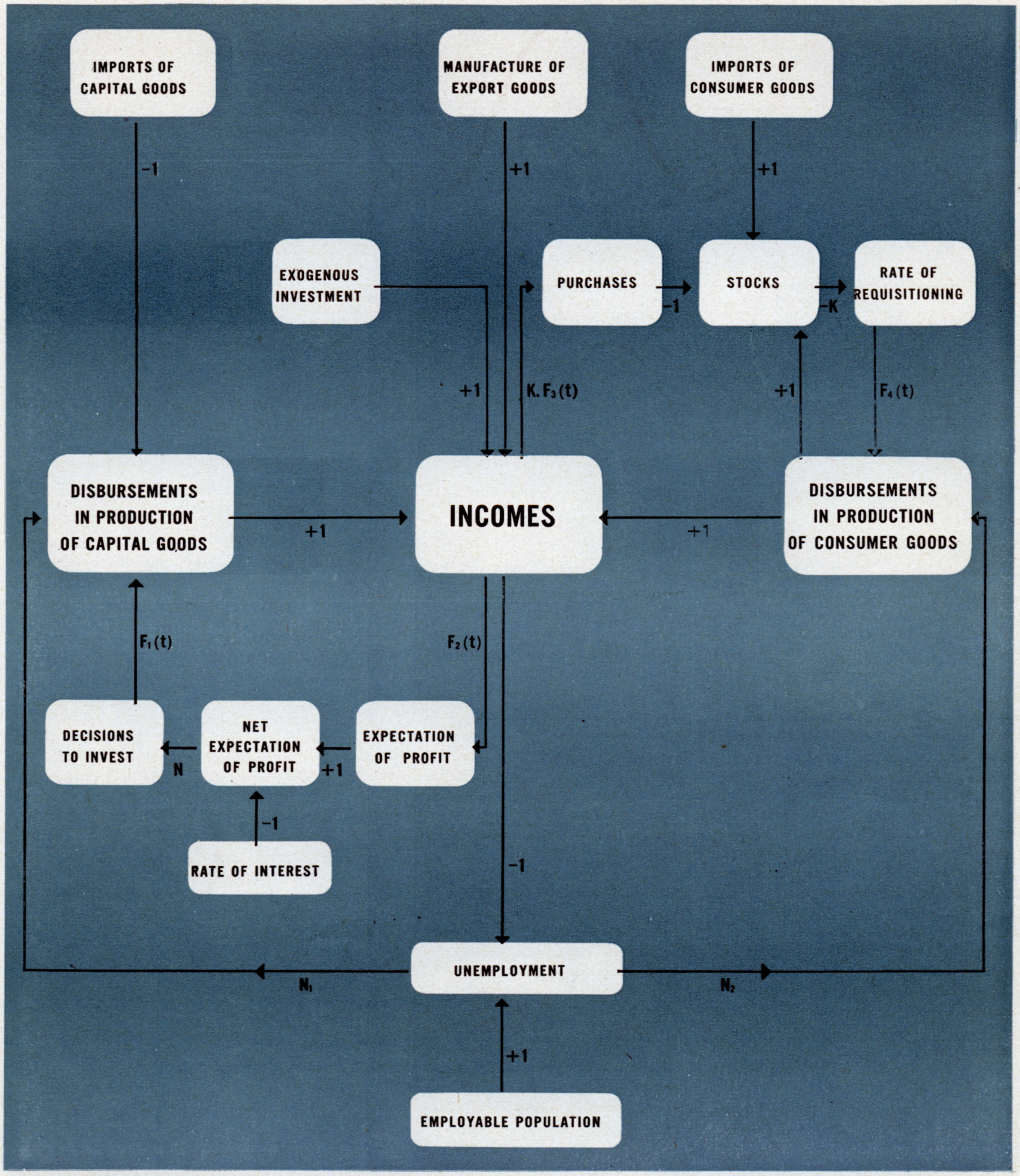
Whatever the nature of the means or medium involved, the function of the central nervous system in the higher animals is clear. It is to provide a biologically more effective control of behavior under a combination of inner and environmental stimuli. An inner analogue or simulation of relevant aspects of the external world, which we are aware of as our idea of the environment, controls our responses, superseding mere instinct or reflex reaction. The world is still with us when we shut our eyes, and we use the "play of ideas" to predict the consequences of action. Thus our activity is adjusted more elaborately and advantageously to the circumstances in which we find ourselves.

This situation is strikingly similar in principle (though immensely more complex) to the introduction of a predictor in the control of a gun, for all predictors are essentially analogues of the external situation. The function of mind is to predict, and to adjust behavior accordingly. It operates like an analogue computer fed by sensory clues.

It is not surprising, therefore, that man sees the external world in terms of cause and effect. The distinction is largely subjective. "Cause" is what might conceivably be manipulated. "Effect" is what might conceivably be purposed.

Man is far from understanding himself, but it may turn out that his understanding of automatic control is one small further step toward that end.

Arnold Tustin is professor of electrical engineering at the University of Birmingham.



FEEDBACK IN AN ECONOMIC SYSTEM is blocked out in this diagram by the author. The scheme of dependence is based upon the ideas advanced by the late J. M. Keynes. Total incomes arise from disbursements for consumer goods on the one hand and capital goods on the other. But each of these is dependent in turn, *via* its subsidiary closed loop, upon total incomes. Keynes was especially concerned with the factors which determine the relationship between the two loops and the relative flow of money into them from total incomes. He showed this to be a highly sensitive relationship, since a comparatively small increase in the flow of money around the capital-goods loop is amplified *via* the consumer-goods loop into a much larger change in total incomes. This is precisely analogous to the behavior of similarly coupled electrical feedback circuits. Pursuing the analogy, the author has entered on the diagram some symbols for values that would have to be defined to design a complete electrical analogue for the economic system. K, for example, is Keynes' "propensity to consume"; F_1 (t) represents the time-lag between the decision to invest and the purchase of capital goods; N_1 and N_2 stand for non-linear functions which curtail increase in production as unemployment approaches zero.