A CRITIQUE OF CYBERNETICS
Author(s): HANS JONAS
Source: *Social Research*, Vol. 20, No. 2 (SUMMER 1953), pp. 172-192
Published by: The New School
Stable URL: http://www.jstor.org/stable/40969483
Accessed: 28/06/2014 10:33

A CRITIQUE OF CYBERNETICS*

BY HANS JONAS

In 1782 James Watt patented the flyball governor for his steam engine. It consists of two balls linked to a vertical spindle which is rotated by the engine; their weight, tending to keep them down and close to the spindle, is counteracted by the centrifugal force of the rotation. With an increase in speed this force causes the balls to fly out and up, while a decrease makes them fall. In thus moving, they operate a lever so connected with the throttle-valve between the boiler and the engine as to close it when the speed exceeds a normal value and to open it when the speed falls short of that value. The beauty of this self-regulation is in the fact that the machine performs it as part of the output to be controlled, and through the very acts of excess or deficiency which are the objects of the corrective action.

Note here the two important aspects of this control mechanism. First, a part of the output energy, though of insignificant magnitude, is redirected to the controlling apparatus farther back in the causal order of the system; this feature is called "feedback." Second, this feedback is such as to counteract the action of the machine, that is, it is corrective, not reinforcing; this is called "negative feedback." When properly functioning it will keep the performance around a mean value by reacting alternately to the "plus" or "minus" departures from it.

More than eighty years later, in 1868, Clark Maxwell, in a paper "On Governors," read before the Royal Society, gave the first theoretical account of this type of mechanism. And again eighty years later, in 1948, Norbert Wiener of the Massachusetts Institute of Technology lifted from the font a new science which he

* EDITORS' NOTE—This essay is based on a paper read before the General Seminar of the Graduate Faculty of the New School for Social Research, on the evening of January 14, 1953.

christened cybernetics, taking the name from the Greek word *kybernetes*—helmsman, pilot—of which our "governor" is a derivative.

I

Little did Watt dream of these consequences. His governor was an auxiliary device for his steam engine, the purpose of which was the production of mechanical power for industry. From the plentiful availability of such power stemmed the industrial revolution, or what Wiener prefers to call the first industrial revolution. Its dominant technological aspect was power engineering. The pilot function of the governor was confined to assuring the steady running of the power engine, and anything that Watt or his contemporaries may have been able to foresee of his invention's effects was surely in terms of the moving force generated by the new machines, and its impact on economy—that is, in terms of the first industrial revolution.

Lately, however, automatic pilot devices have come into their own with a difference. Modern technology, going beyond the mere production and application of power, tends increasingly to couple the power engine with robot mechanisms—mechanisms that replace man's perception and judgment in the serving of the machine, just as the power engines replaced man's arms in supplying the moving force. The difference lies not only in the function but also in the technology; automatic control is a branch of communication engineering, as distinct from power engineering. It is the rise of these servo-mechanisms, and the fact that they supersede human functions very different from those superseded by the mere power engine—generally speaking, "higher" functions—which causes Wiener and others to speak of a second industrial revolution. Familiar examples of servo-mechanisms are the thermostat, self-correcting steering engines in ships, automatic fire control in anti-aircraft artillery, target-seeking torpedoes, electronic computers, automatic telephone exchanges. In all of them, feedback plays an important part.

That the common principles involved in these different devices and the problems posed by them require a unified theory, and that this theory is of sufficient autonomy and range to deserve the name of a new science, was a matter for the workers in the field to decide, and no quarrel can arise with their practical decision. If this were all the status claimed by cybernetics, the philosopher would not have to engage in a critique of it.

But cybernetics is not as innocent as that. There is a strong and, it seems, almost irresistible tendency in the human mind to interpret human functions in terms of the artifacts that take their place, and artifacts in terms of the replaced human functions. The power engine, with its levers and joints and its voracious fuel consumption, was a slaving giant, and, correspondingly, the human or animal body was a fuel-burning power machine. The modern servo-mechanism is described as perceptive, responsive, adaptive, purposive, retentive, learning, decision-making, intelligent, and sometimes even emotional (but this last only if something goes wrong!), and, correspondingly, men and human societies are being conceived of and explained as feedback mechanisms, communication systems, and computing machines. The use of an intentionally ambiguous and metaphorical terminology facilitates this transfer back and forth between the artifact and its maker. In former days, dealing in such analogies was left mainly to the imaginative writer, and certainly had no part in the terms of reference of the scientist as such, but this sort of transference is precisely what cybernetics is concerned with, and on this account it is subject to philosophical criticism. The literature, which fortunately is not yet too bulky, already abounds with cybernetical explanations of human behavior, processes of thought, and cultural organisms.

This is something new. The classical mechanist, dealing with matter and motion, was content to speak of the "machine of the body" and, in method if not in metaphysics, to follow Descartes, originator of the notion of the "animal automaton," who had removed mind from the very terms of physical science. Later

materialists repudiated the Cartesian dualism in name only. That mind is an epiphenomenon of material processes in the brain, as they held, remained a summary assertion not leading to such actual correlations and transformations as would invade the field of mind itself with the symbolism of physical science. Now is offered, for the first time, a mechanistic model which, it is claimed, applies to material and mental phenomena at once, not only equivalently but identically—that is, without involving passage from field to field.

This would indeed mean an overcoming of the dualism which classical materialism had left in possession by default: for the first time since Aristotelianism we would have a unified doctrine, or at least a unified conceptual scheme, for the representation of reality. Needless to say, this would be of the utmost philosophical importance—and the spokesmen of cybernetics are not restrained by any timidity from pointing out these implications in very explicit statements. It is with this aspect of the new discipline, not with the mathematics and technology of communication engineering and automatic control, that this paper proposes to deal.

There are three major topics facing such a scrutiny, indicated by the terms "teleology," "information," and "mind." I have chosen for analysis here the first of them: the cybernetical concept of purpose and teleology. This is basic to the whole scheme, and I grant at the outset that if cybernetics makes good its claim with regard to these terms—that they can be evolved from mechanical premises alone—it has carried its main point and resolved an age-old dualism. I propose to show that the resolution claimed is spurious and mainly verbal.

Before entering into detailed discussion, let me show by a stock example from cybernetical literature what kind of analogies we are about to deal with in this field.[1] The analogy in question is

[1] The example is taken from a joint paper by A. Rosenblueth, N. Wiener, and J. Bigelow, "Behavior, Purpose and Teleology," in *Philosophy of Science,* vol. 10, no. 1 (January 1943) pp. 18–24; this paper is the source of all subsequent quotations, except where otherwise indicated.

between a servo-mechanical and a neurological disturbance. This is the mechanical side: a feedback may be inadequately "damped" and result in overcorrections, thereby in effect becoming positive instead of negative; in such a case a machine, say one that is designed to hit a moving target, will "overshoot" in alternate directions in a series of ever larger oscillations, and thus "miss the goal." And now the neurological analogue: if a cerebellar patient "is asked to carry a glass of water from a table to his mouth . . . the hand carrying the glass will execute a series of oscillatory motions of increasing amplitude as the glass approaches his mouth, so that the water will spill and the purpose will not be fulfilled." The condition described is called "purpose tremor."

It is contended that the two cases are "strikingly similar." However that may be, one point at least should be clear from the outset: the patient himself *wills* to bring the glass to his mouth, that is, he wants it there. This *end,* motivating the action from the start, is intrinsic in all the part-motions, providing the reference by which they are in themselves failures and make the whole undertaking a failure. Presumably the patient finds his inability to perform distressing. But the machine, for all we know, may just as well be said, instead of being distressed, to abandon itself with relish to its wild oscillations, and instead of suffering the frustration of failure, to enjoy the unchecked fulfilment of its impulses.

"Just as well" amounts of course to "neither." Manifestly neither "distress" nor "enjoyment" fits the modus operandi of a machine—not even as an operational analogy, since the machine is equally "satisfied" in each and any single step of behavior as it occurs, the occurrence as such being its own sole and sufficient vindication. In the case of the machine "missing *the* goal" means, of course, missing *our* goal, the goal for which it has been designed, namely by us, it "having" none itself; whereas the unfortunate patient truly misses *his* goal, which is his not because he has been designed for it but because he has formed and entertained the design. These elementary distinctions, for whose probably exas-

perating obviousness I apologize, have to be kept firmly in mind during the following deliberations. The example itself, however, belongs to a later stage in the logical development, to which I now turn.

II

In the cybernetical study referred to, the concept of purpose emerges in a process of dichotomy to which the basic concept of "behavior" is subjected. Since "by behavior is meant any change of an entity with respect to its surroundings," the whole sequence remains within the terms of external relationships. Purposeful behavior appears as a subdivision of active behavior, and "the term purposeful is meant to denote that the act or behavior may be interpreted as directed to the attainment of a goal—i.e. to a *final condition* in which the behaving object reaches a *definite correlation* in time or in space with respect to another object or event" (italics mine).

Obviously the whole definition turns on the meaning and relevance of the term "final condition." Now what condition is to be regarded as final? We are not permitted to answer "the one in which the goal is reached," since it is the finality of the condition which alone gives meaning to the term "goal"; this meaning is not derived, for example, from the anticipatory presence of such a condition in the initiation and throughout the successive stages of the motion. Nothing remains but to understand as "final" the condition in which the action ends, that is, a condition of rest, in the broad relativistic sense indicated by the phrase "in which the behaving object reaches a definite correlation . . . with respect to another object." This sounds almost Aristotelian, except that to Aristotle a body comes to rest in its natural place because this *is* the aim of its motion, while to our authors the motion "may be interpreted" as having had that aim because it ends where it does.

The difference is not a minor one. Aristotle could distinguish between the mere ending and the intrinsic "end" of a motion, a

distinction without which, as he points out, death would have to be considered the aim of human life. But to our authors, if they stick to their definition, death—the most definite correlation to the environment reachable by an organism—indeed *is* the goal of the total motion of life as one sequence of "active behavior," and I see no way for them to escape this conclusion. To generalize this, we may say that "running down," that is, increase of entropy, defines the direction of all natural processes, and therefore maximum entropy is the goal to whose attainment all behavior may be interpreted as being directed. In this sense all behavior is purposeful—by the terms of the definition.

This criticism may seem unfair, since it does not consider the kind of mechanism to which the authors intend their definition to apply, and which, by its statable difference from other "non-purposeful" kinds, exemplifies the true meaning of the definition. Indeed, we are not concerned with verbal shortcomings. Let us therefore look at some of the examples given to illustrate the meaning—first at a non-purposeful mechanism.

"First may be mentioned mechanical devices such as a roulette, designed precisely for purposelessness." What a mechanism is designed for, by its maker, is of course entirely irrelevant, because extraneous, to the description of its working. The mention of the purpose of the human designer in this context arouses our suspicion that also in the general theory the human—that is, the familiar, the non-cybernetical—meaning of "purpose" is insinuated into a description which purports to deal in terms of external behavior alone. We shall find this suspicion amply confirmed as we go on. As regards the roulette itself, it attains in each run one final condition, though one that is unpredictable by its designer or user, and is thus by itself "purposeful." It may be argued that it does not come under a definition by which "purposeful" is a subdivision of "active" behavior, since the whole action of the roulette can be traced to the immediate input of energy, the push by a hand. But instead of losing time with showing that the distinction between active and passive behavior is

irrelevant to the issue, let us go on to the next example, which is not open to this objection.

"Then may be considered devices such as a clock, designed, it is true, with a purpose, but having a performance which, although orderly, is not purposeful—i.e., there is no specific final condition toward which the movement of the clock strives." This is a glaring example of illicit mixing of points of view. The statement that there is no specific final condition toward which the movement of the clock strives is true only as an expression of desirability from the point of view of the human user, who intervenes—by periodical rewinding, for example—to secure this desirable state. Left to itself, to its own "striving," the clock will run down: the spring will extend to its maximum, the zero point of tension, or the weight will come to rest, at the latest on reaching the center of gravitation, and the final condition of the clock itself will be that standstill which marks the attainment of complete equilibrium or maximum entropy.

I cannot see any other striving in the clock as a piece of physical mechanism, though I can see a striving of a very different nature in the *user* of the clock to counteract the clock's own "purpose." On the other hand, it seems arbitrary to exclude the regular performance as a whole from the conditions "in which the behaving object reaches a definite correlation in time or in space with respect to another object or event," and thus to disqualify it from representing the goal of the process. The "definite correlation" may well be a series, which it is in the case of a continuously energized clock (and, on the human level, in the case of all those activities that have their end in themselves); and thus in this respect too the example fails to illustrate non-purposive as opposed to purposive behavior in a mechanism. It seems that once we have abandoned the original meaning of "purpose" as the *propositum,* that which someone sets *before* himself as the *whereto* of his action, we are reduced to the necessity of granting purpose to *all* action—thereby depriving the definition of all defining force.

Let us now look at our authors' counter-examples. "Some machines, on the other hand, are intrinsically purposeful. A torpedo with a target-seeking mechanism is an example. The term servo-mechanisms has been coined precisely to designate machines with intrinsic purposeful behavior." Why is a target-seeking or self-steering torpedo intrinsically purposeful? We must of course beware of succumbing to the suggestiveness of such words as "seeking" and "self." In observing the behavior of the torpedo, its "change with respect to its surroundings," we notice that it does not simply follow its initial course, but alters it occasionally, with the general effect of keeping on the target, which may be a moving one. We see these changes occur in response to changes in the relative positions of the two entities, and may then speak of "compensatory action" on the part of the torpedo. We do this provided the changes of behavior are due not directly, in terms of energy, but only indirectly, in a sense to be specified, to the influence of the other entity. Otherwise any magnetic particle changing its path with the displacement of a magnetic pole, or a gravitating mass changing it with the relative displacement of the center of gravitation, would have to be regarded as coming into the same category.

The latter example is instructive. A planet may be said to respond continuously to the relatively changing position of the sun, and the curvature of its actual path may be said to be a compromise between two conflicting "strivings," one to continue in a straight line by its momentum, the other to fall toward the sun by its gravity. Each factor embodies its own "purpose," and if the first becomes sufficiently small the revolution of the planet will become a decreasing spiral in which the margin of "error" with regard to the second "purpose" will be progressively diminished. Yet we would not say that the planet "corrects" or "adjusts" its course. The reason is that here the changes are a direct function of the forces involved; especially, the energy that provides the "stimulus" (the "pull" of the sun) is the same as that which effects the "response." In other words, we cannot here even make

the distinction, since action and reaction are one and the same event.

An analogous situation would obtain if the torpedo altered its course by direct magnetic attraction between itself and the target. This would certainly be "purposive" behavior according to any reasonable interpretation of the previous definition, which has been shown to be worthless in any case, but it would not be "teleological" behavior according to the definition we are soon to deal with. What constitutes the difference in the two cases, assuming that magnetic principles operate in both, is that in the self-steering torpedo the magnetic element has no part in the *acceleration* of the entity whose steering arrangement it affects, and the effect on the latter is not a function of the quantity of the magnetic force acting on it. Given sufficient sensitivity, this force may be as small as you please, and given efficient coupling and sufficient motor resources, the effect in terms of power may be as large as you please. The torpedo is not attracted but is steered toward the target—in response, to be sure, to an influence emanating from it, but this influence is of the order of "message" and not of acceleration. Thus it is the difference between receptor and effector elements, and the steering of the latter's output by the former's input, which characterize this type of adaptive behavior. On the human level this is tantamount to saying that purposive behavior involves perception.

Obviously the division, and at the same time connection, between receptor and effector organs is one of the essential conditions of the freedom of animal action: it makes possible action governed by purposes, provided the entity in question is one that can have purposes. This instrumental condition, which is no other than the "feedback" of the control engineers, provides the cybernetician with his definition of "teleological behavior": behavior controlled by negative feedback. Thus it is a subdivision of "purposive behavior," and it escapes the meaninglessness of the generic term through being thus specified by a definite technical pattern.

With negative feedback an entity's function is controlled by the amount of discrepancy which it shows from moment to moment with reference to some defined state, the amount being continually compensated. The process thus appears to be goal-directed. Can it therefore be called "teleological" in any more relevant sense than that of a verbal definition—in a sense that would justify the choice of the name from the connotations it has outside the definition? The question is tantamount to asking whether the differentia of "feedback" can supply the concept of *purpose* which the logical genus failed to yield; this again amounts to the question whether the technical *condition* for purposive action can itself constitute purpose; and this, finally, involves the question whether effector and receptor equipment—that is, motility and perception alone—is sufficient to make up motivated animal behavior.

III

With these questions in mind let us look again at the target-seeking torpedo. Its *telos* would be said to be the hitting of the target, because this will be the result of its behavior if successful, and the behavior is said to be purposeful and teleological because it is self-adaptive with regard to this result. Note that in the term "successful" we have introduced an element of human evaluation into the description, and that in saying "adaptive with regard to the result" we have allowed ourselves another anthropomorphic latitude, since there is nothing in the single operations of adaptation that relates directly to the outcome of the action as a whole, even though retrospectively we can adjudge the single adaptations to have contributed to it. Each of the adaptations in itself only restores the equilibrium of the moment, which on its own terms is a self-sufficient situation. I will not press these points here, but rather ask what *part* of the mechanism embodies the purposiveness.

It cannot be the propulsion engine, because this—whether it is a battery-fed electromotor, a jet or internal-combustion engine, or whatever—simply works toward the equalization of energy levels,

that is, it is governed by the law of entropy. But so too is the sensitive receptor, whose action from moment to moment consists in the leveling-out of a magnetic or electric potential or some other internal disequilibrium; and the same is true for the transmission of its "message" through whatever channel, and for its amplification, and for the relay actions involved, and so on, back to the operation of the steering device as such. Each of these actions runs its course entirely in the tracks of its own mechanical necessity—"blindly," as the expression is—and is as unrelated to the previous and following steps in its own series as to the actions of the other elements in the system. None of them is as such engaged in the attainment of the "goal," the sole "concern" of each being the attainment of its own "purpose" in terms of entropy.

Thus the overall purpose, since it does not reside in any part, must reside in the whole, in the receptor plus the effector plus the coupling, and in the form of organization of the multiple system. This indeed is the cybernetical contention, and therefore the question becomes that of whether the mechanism *is* a "whole," having an identity or selfness that can be said to be the bearer of purpose, the subject of action, and the maker of decisions.

That it is not can be shown by a simple mental experiment. Imagine the torpedo, not fully mechanized, to be manned by a human pilot—and of course you may immediately substitute for this image the everyday example of the driver in his car. One would no more regard the torpedo plus the brave soldier, or the car plus the driver, as a single purposive entity than one would declare the axe to participate in the purposiveness and the teleological behavior of the lumberjack who swings it. Any sane person would say that the soldier, the driver, the lumberjack, is the bearer and agent of purpose in the combination, and uses its other elements to his purpose. And this situation is not altered at all if the pilot, for example, has no first-hand perception of the goal but does his steering by the data of mechanical receptor devices such as radar. In this case we should have mechanical

feedback input and mechanical energy output coupled by a human link at the control point, but this still would not bring the machine a jot nearer to merging with the human agent into one purposive whole. The man can step out of the contrivance and walk away and take the "purpose" with him, complete and unabridged.

In certain cases, therefore, we *can,* within an overwhelmingly mechanical system of interlocking functions, including even servomechanisms, *localize purpose* in one single controlling part—provided this part or link is such an entity as for itself *has* purpose and *acts on* purpose. All the rest of the machinery is then simply his tool. Why, then, if we now replace this one agent by a mechanical link, which taken by itself has no purpose, should the quality of purposiveness shift from this locus in the configuration and suddenly expand over the whole of the system, and the former tools assume, together with the replacing element, the characteristics of intrinsic purposeful behavior? The idea is absurd. We may say, of course, that the whole contrivance, judged from without, behaves "as if" it were purposeful, only we must immediately add "but we know better."

Why do we know better? To answer this question let us concentrate for a moment on the hypothetical human pilot of the target-seeking torpedo. He carries out certain actions, performs certain motions—those that would be carried out in the alternative case by the interpolated mechanical device. Why does he perform those motions as and when he does? The cybernetical answer would be that he does so in response to certain information from his receptors, or he acts as determined by the latter—in other words, he functions as a feedback mechanism himself.

But this is patently untrue. Everyone in his right senses will say that the pilot operates a certain switch not because he has received a certain sense-message but because he wishes to keep the torpedo on the target, and *in the light of this purpose* he takes the occurrence of certain perceptions as the occasion for performing certain actions conducive to the end in view. It is the prior and coextensive purpose which qualifies the incoming data as messages if

they are relevant to the purpose. It is I who let certain "messages" count as "information," and as such make them influence my action. The mere feedback from sense-organs does not motivate behavior; in other words, sentience and motility alone are not enough for purposive action—not even for the original conditioning of reflexes which, once set up, may then substitute for purposive action. The reflex arc embodies in its mechanized pattern the vital purposiveness or concern that went into the making of it. The feedback combination of a receptor-effector system lends itself to purposive action precisely if and when it is not a feedback *mechanism*—that is, if the two elements are not coupled directly, but if interposed between them there is will or interest or concern.

This amounts precisely to saying that purposive behavior requires the presence of purpose. That statement is no mere tautology, for cybernetics is an attempt to account for purposive behavior without purpose, just as behaviorism is an attempt at a psychology without the "psyche," and mechanistic biology a description of organic processes without "life."

If then everything turns on purpose itself, let us eye more closely the purpose of our torpedo pilot. We said he acts so and so because he wishes to keep the torpedo on a certain course. But the matter does not end here. This being only the most proximate purpose, we immediately have to ask why he wishes the torpedo to follow this course.

One answer might be that he wishes the torpedo finally to hit and sink the enemy cruiser. And why should he wish that? This is all the more relevant to ask, as the desired success may entail his own destruction, and we may credit the agent with a concurrent desire to live. The answer may be that he wishes it out of patriotic fervor, or for honor and glory, or out of hate or revenge, or to win a bet, or to commit a spectacular suicide. But in each of these cases, or in a combination of them, we again have to ask why, and thus launch into a consideration of what the welfare of his nation, or honor or glory or revenge or whatever, means in the total teleological economy of his person. This investigation will

lead to his ultimate concerns, the ends by which he lives, at least at the present juncture.

But the first answer might have been not "because he wishes to sink the cruiser," but "because he has been ordered to steer the torpedo to that effect," and then the next question would be "why does he make obedience to orders his purpose in this sequence of action?" Here the answers might be that he does so out of a sense of duty, or out of fear of his superiors, or to win their approval, or because he wants to earn his pay, or to conform to a social pattern. Each of these answers leads again into an ascending and expanding series of mediate purposes or concerns, and when followed through will end up in a picture of the total purposiveness of the man.

The point is that, however limited and possibly heteronomous the immediate motivation may be, it can become a motivation only on the basis of the concernedness of all life with itself, its performance, its content. Only on this basis can the "feedback" operate in the control of purposive action. Even the appeal to obedience and the widest use of habit must ultimately draw on this fund of spontaneity and interest.

But our second set of hypothetical answers is instructive in a further respect. We assumed that the purpose once removed from the immediate purpose of "keeping on the target" was "carrying out orders." Although this, as we have seen, will itself fall into the pattern of the agent's own overall purposiveness, it points also to somebody's else purpose, which the agent can be said to be carrying out with his action. This need not mean that he has made the commander's purpose his own, but he certainly has made its execution, as far as entrusted to him, his present purpose, and this for purposes of his own, as was pointed out.

From the point of view of the commander, however, these purposes of the pilot do not matter at all. All he is interested in is that he can reasonably rely on the agent to execute his orders, for whatever reasons, so that in his own pragmatic account he can substitute for those reasons his own orders as the sole effective

determination of the other's behavior. He can do so, of course, only if he knows that the agent's own purposiveness includes the accepting of orders from him—and of such orders as the one given. But once this condition is observed, as it will be by any intelligent commander, he may indeed forget about the other's motives. In other words, he can regard the pilot during his mission as his robot, his tool.

Seen in this way the pilot indeed merges with the torpedo into one instrumental entity, and the whole represents a servo-mechanism. But this means that the combination has *no intrinsic* purposefulness in that consideration, and merely carries the purpose of its user. And this is the very case of the pilotless "self-steering" torpedo. The human pilot *can* be replaced by a mechanical device precisely because his own intrinsic purposiveness does not count in this context; what counts is only the purpose of the commander, who by the "remote control" of his orders, or by the instructions "fed in" at the outset, operates the "machine"—propulsion engine, pilot, and all.

Thus there is indeed a sense in which we can say that a purposive action is being performed by the whole complex, whether with or without a human pilot in its combination. We still remain in agreement with the stated axiom that purposive behavior requires the presence of purpose, if the purpose here present is understood to be that of the commander, that is, a purpose extrinsic to the system. In the absence of that, the mechanism, even with identical action, becomes purposeless on this level of consideration—though by being active it inevitably performs the "purpose" intrinsic to all mechanical action, the attainment of entropy; or, if it includes a human element, this will assert its own purposiveness and perhaps steer the machine and itself to safety.

IV

In terms of mere semantics we may say that the whole cybernetic doctrine of teleological behavior is reducible to a confusion of

"serving a purpose" with "having purpose." Similar statements would apply, mutatis mutandis, to the cybernetic concepts of "information" and "reasoning." But semantic confusion does not explain the phenomenon; it serves only to diagnose it.

We come nearer to an understanding if we realize that the cybernetician looks at his objects in a theoretical situation somewhat like the practical situation in which our commander looks at his subordinate—that is, a situation in which indeed the distinction between man and machine is irrelevant and the two become interchangeable. But the commander, though in his handling of the situation he takes the subordinate as a robot, does not take himself to be one—and this in spite of his being well aware that in the action context of his superiors he in turn is to them a robot, instrumental to their purposes—and so on. His knowledge that he is thus viewed from without, and that he is always capable of being thus viewed, does not cast doubt on the knowledge he has of himself from within. Reflecting on this, if he has the time, he will apply the same consideration to his subordinate, and grant him that he is of course not really a robot.

It is this reflection which the cybernetician fails to perform. He himself does not come under the terms of his doctrine. He considers behavior, except his own; purposiveness, except his own; thinking, except his own. He views from without, withholding from his objects the privileges of his own reflective position. If asked why he embraces cybernetics, he would for once answer not in cybernetical terms of feedback, circular loops, and automatic control, but in terms like these: "because I think it to be true, and I am interested in truth"; or "because I think it to be useful for such and such ends, and I am interested in those ends"; or "because it is the rising fashion, and I like to keep up with the times"; or whatever else may be truthfully or untruthfully answered in such cases.

But if asked why a group of persons other than himself organize a conference on cybernetics, he would answer that there are "many regenerative loops" in the single nervous systems which, in

their circular paths, perpetuate signals as "universals," and that "by joining these loops universals can be related" and "thereby the postulates of any . . . theory . . . constructed"; and that if such a "related system of impulses in reverberating circuits . . . gets into a nervous system so as to define the form of its activity [it may] determine the pattern of firing of motor neurons, and so literally, causally and neurologically determine an overt, objective, social and institutional fact." [2] Nothing could be more devastating for this account of theory-forming than to be found self-illustrative. Professor Northrop would be justly indignant were I to suggest that this theory of his has no other logical status than that derivable from the kind of genesis it describes, and were I, in his own case, to substitute the process there propounded for his seeking after and conforming to truth.

We are here, as in so many other cases, in the presence of what I would call split-personality theorizing—a phenomenon unavoidable, and to that extent excusable, in some of the special sciences, but inadmissible and fatal in philosophy, and hardly less so in those sciences that include man among their objects. In abstracto the behaviorist must count himself among the objects of his method. But in concreto he must make the implicit reservation of self-exemption, at least with regard to his reasoning in support of the behavioristic thesis, for the sake of its claim to validity. Furthermore, expecting his argument to be evaluated on its merits, he must also exempt those to whose judgment it is addressed in scientific discourse, while at the same time considering them as instances of those "other than myself" to whom the method should apply. And he himself is being considered by them with the same duplicity—inside and outside the discourse.

[2] To quote F. S. C. Northrop, "The Neurological and Behavioristic Psychological Basis of the Ordering of Society by Means of Ideas," in *Science*, vol. 107 (April 23, 1948) pp. 411 ff. Another victim of cybernetics from the ranks of the philosophers is K. W. Deutsch in his article, "Mechanism, Teleology, and Mind," in *Philosophy and Phenomenological Research*, vol. 12, no. 2 (December 1951) pp. 185 ff. It is against philosophical excrescences such as these, much more than against the specialistic work of the cyberneticians proper, that the present criticism is directed.

If he reflects sufficiently he may even be aware of all this. The same necessary duplicity obtains in the case of the materialistic biologist, the same in the case of the cybernetician.

These cases need not too much disturb us, since the special sciences, after all, are concerned not with the whole but with isolated aspects that they tackle in their terms at the declared cost of unity. But what are we to say of philosophers who, taking their cue from one or another of the special sciences, cast themselves with the abandon of self-abnegation into the disavowal of the ego cogitans?

This is a subject too wide to broach here, sadly as it needs broaching. But in the present case cybernetics cannot be entirely absolved from guilt. It is not the innocent special science which seduces susceptible philosophy by its passive beauty: from its inception it has been out to capture her. From its inception it has pretended to the status of a unified theory of mechanism, organism, the nervous system, society, culture, and mind; and by its suggestive employment of the terms behavior, purpose, goal, information, memory, decision, learning, initiative, value, and thought it has so inflated its initially modest definitions that their resulting use amounts to hardly more than verbal trickery.

V

In conclusion, and for an observation transcending the mere negative criticism with which this paper has unfortunately had to be filled, let me single out one level—the biological—from all the levels of reality to which cybernetics applies itself. Until corrected on this point by more competent experts, this layman is ready to concede that the sensor-effector combination in animals does in certain respects represent a feedback pattern, and, to the extent that it does, conforms to the model evolved by cybernetics. Where, then, does that model fall short?

The answer can be compressed into one statement: living things are creatures of need. Only living things have needs and act on needs. Need is based both on the necessity for the continuous

self-renewal of the organism by the metabolic process, and on the organism's elemental urge thus precariously to continue itself. This basic self-concern of all life, in which necessity and will are bound together, manifests itself on the level of animality as appetite, fear, and all the rest of the emotions. The pang of hunger, the passion of the chase, the fury of combat, the anguish of flight, the lure of love—these, and not the data transmitted by the receptors, imbue objects with the character of goals, negative or positive, and make behavior purposive. The element of effort alone lifts bodily activity out of the class of mechanical performance, and the fact that movement requires effort means that an animal will move only under the incentive of an interest.

The cybernetical model reduces animal nature to the two terms of sentience and motility, while in fact it is constituted by the triad of perception, motility, and emotion. Emotion, more basic than the two it binds together, is the animal translation of the fundamental drive which, even on the undifferentiated pre-animal level, operates in the ceaseless carrying on of the metabolism. A feedback mechanism may be going, or may be at rest: in either state the machine exists. The organism has to keep going, because to be going is its very existence—which is revocable—and, threatened with extinction, it is concerned in existing. There is no analogue in the machine to the instinct of self-preservation—only to the latter's antithesis, the final entropy of death.

According to cybernetics, society is a communication network for the transmitting, exchanging, and pooling of information, and it is this that holds it together. No emptier notion of society has ever been propounded. Nothing is said on what the information is about, and why it should be relevant to have it. The scheme allows no room for such a question even to be raised. Any theory of man's sociability, however crude or distorted, that takes into account his being a creature of need and desire, and that looks for the vital concerns which bring men together, is more to the point. Grim old Hobbes showed himself infinitely better informed than the information specialists when he contended that

fear of violent death and the need for peace brought men into the covenant of commonwealth and continue to hold the body politic together. Onesided as his doctrine may be, it is valid in that it ascribes man's action to a striving after some good, even if this be the mere preservation of life. Without the concept of good, one cannot even begin to approach the subject of behavior. Whether individual or social, intentional action is directed toward a good. According to some, the scale of lesser and greater goods that can become the objects of desire, and thus motivate behavior, culminates in a highest good, the *summum bonum*. In the case of man this may well be, in a sense very different from that of cybernetics, *in-formation*.