# The Neglected Phenomenon of Negative Feedback Control

William T. Powers

## INTRODUCTION

In a commentary in *Nature*, Rodney Brooks[1] proposed that something is missing from our models of living and behaving systems. I would like to suggest that it is not something undiscovered that is missing, but something old that has been passed over without sufficient examination. What is missing from most "modern" conceptions of behavioral mechanisms may be a sufficient understanding of a remarkable phenomenon called negative feedback control, reduced to a formal theory over half a century ago.

Systems organized to carry out negative feedback control behave in a way that a great many scientists do not believe is possible. Given a specification for some state of affairs, they can continue to produce or reproduce the specified outcome even though the actions needed to do so vary from one moment to another. The actions of such systems are of the type that has been termed *purposive*, in that they appear designed to achieve some specific predetermined end. They are also of the type that has been termed *adaptive*, for such systems are able (within limits) to vary their actions in just the way needed to continue to produce a particular outcome despite changes in circumstances.

In contrast, what most life scientists seem to believe in can be termed a *causal* system. A causal system mediates, stands between, causes and effects. The effects created by a causal system are those dictated by its physical structure and external forces or other influences acting on that physical structure. If circumstances change, the effects necessarily change, either because the behavior-causing external forces and influences change or because the structure of the system is changed by other forces and influences. What we see a causal system doing corresponds to what is being done to it; its "actions" are more properly called "responses," for no action of a causal system takes place without an adequate prior external cause or stimulus.

Before the 20th century was half done, engineers had discovered (and rediscovered) the phenomenon of negative feedback control and had founded a new formal discipline, control engineering. But this new concept clashed with what most scientists concerned with living systems already believed. From the very start there was a concerted attempt to assimilate the new concepts of control into the old ideas of causation.

The result has often been a strange blending of purpose and causation – for example, the frequently-used idea of an organism learning how to respond the most effectively to stimuli or "cues" from the environment. The idea of responding to cues or stimuli belongs in the causal model, but to "respond the most effectively" requires the organism to perceive the effects of its own actions and modify the actions so as to achieve some desired degree of effectiveness —a concept that is more appropriate to a negative feedback control system.

Another effect of this blending has been to conceal the problem of purpose by hiding it behind a screen of causal complexity. Brooks (op cit), for example, describes a "behavior-based" approach. "... this new mode of thought," he says, "involves the connection of perception to action with little in the way of intervening representational systems. ... this approach relies on the correct short, fast connections being present between sensory and motor modules." But "correct" implies "correct for achieving a specified outcome," which is a concept that derives from the properties of negative feedback, not simple input-output causation.

Probably the most elaborate blend has come to be called (somewhat hubristically) "modern control theory." As an approach to engineering control problems it has its merits, but as a model of organisms it only reinforces the old causal model. The basic idea is that the behaving organism picks (somehow)

an outcome of behavior that is desired, and then, computing backward through the environment and the acutators with which a control system affects its environment, deduces the quantitative commands that must be issued to create that particular outcome. Once the inverse calculations have been done and the correct commands have been formulated, the system behaves causally, since the commands are converted into actions just as in any cause-effect device. The problem of purpose is put aside by assuming that there is some desirable outcome of behavior, without spelling out what desires or intends it or, for that matter what a desire or an intention is.

These attempts to assimilate control systems into a causal model of organisms have effectively usurped the role of a pure control-system approach, delaying the introduction of negative feedback control concepts into the mainstream of science. At present, the delay amounts to fifty or sixty years, depending on whether one starts counting just before or just after World War Two. There is a backlog of unassimilated evidence from all branches of the life sciences, all the way down to cell biology, that negative feedback control is a basic principle of life processes. Let us review briefly some known systems among the many that have been and eventually will be discovered.

## BIOCHEMISTRY-LEVEL CONTROL

The requirements for making a biochemical negative feedback control system are not complicated. Consider Figure 1, from *The dynamic analysis of enzyme systems* by Hayashi and Sakamoto[2]. The diagram shows a biochemical system in which an enzyme catalyzes the rate of one stage of the main reaction from substrate A through X1 to X4, and in which effects of the last product in the chain are connected back to the enzyme, so that the final stage of the reaction affects a prior stage.

The labels X1 through X4 stand for concentrations of biochemicals, with the arrows indicating reactions that break down one substance to produce another, as in metabolism (not all reaction products are shown; reverse reactions also occur). The Y1 through Y3 labels represent signaling molecules that serve primarily to carry information, being present only in minute amounts. The enzyme in the middle is shown in two states, active ($e_a$) and inactive ($e_i$). When most of the "allosteric" (alternate forms) enzyme molecules are in the active state, they increase the rate at which X3 is used to form X4. When the enzyme molecules are mostly inactive, the rate of the net reaction is slowed almost to zero. Since X4 is being used up all the time through the path $k_4$, the steady-state concentration of X4 is raised and lowered by the activation or inactivation of the enzyme molecules.

The concentration of the signaling molecule Y1 is affected by the concentration of X4. If (the concentration of) X4 increases, (the concentration of) Y1 increases, and the population of enzyme molecules moves more toward the inactive state. But that would decrease the rate of the reaction from X3 to X4 and *lower* (the concentration of) X4, the negative of the change we started with. We can drop the expressions in parentheses if we just remember that, for example, "X4" used to indicate
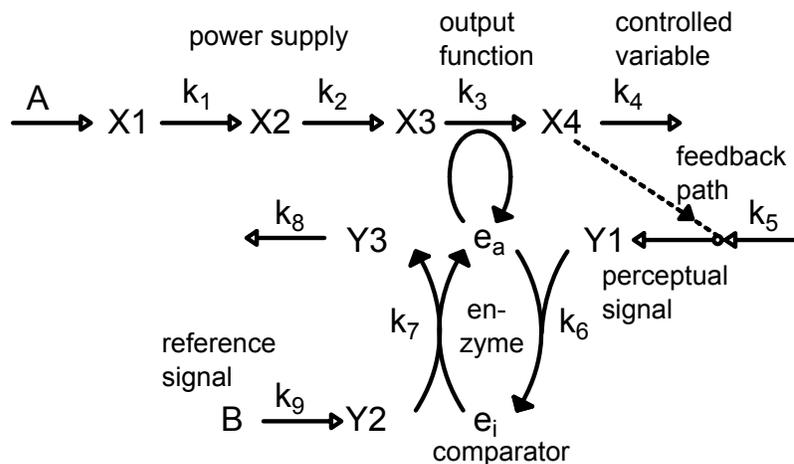


Fig. 1. Biochemical system with annotations suggesting functions in a standard negative feedback control system. X4 is the controlled variable. Redrawn from Hayashi and Sakamoto.[2]

a quantity always means "the concentration of X4". So if X4 increases, the immediate result is for X4 to *decrease* because of feedback effects. This is what is meant by negative feedback.

The state of the enzyme molecules is also affected by another signaling molecule, Y2. An increase in Y2 causes the enzyme population to move toward the active state, increasing the concentration of X4. When X4 increases, however, Y1 increases and progressively inactivates the enzyme. So we have Y2 increasing the enzyme activation, Y2 decreasing it, and X4 being increased by an increase in Y2 and decreased by an increase in Y1, which here is the same as X4.

This may illustrate why the operation of negative feedback control systems has not always been intuitively obvious to a person tracing out the individual relationships in the whole system one at a time. What, in fact, will this circular conglomerate of causes and effects *do* when set free to act by itself? For example, what will happen if we set Y2 to some starting concentration, wait a while, then set it to a different concentration?

The originator of a biochemical simulation program[3] kindly constructed a simulation for the author in which Y2 above was set to one steady level for a period of time, and then halfway through the run was switched to a different steady level. The time scale was such that after each change, the system was allowed to come to a steady state. The result was Figure 2, in which the concentration of X4 is plotted against time.
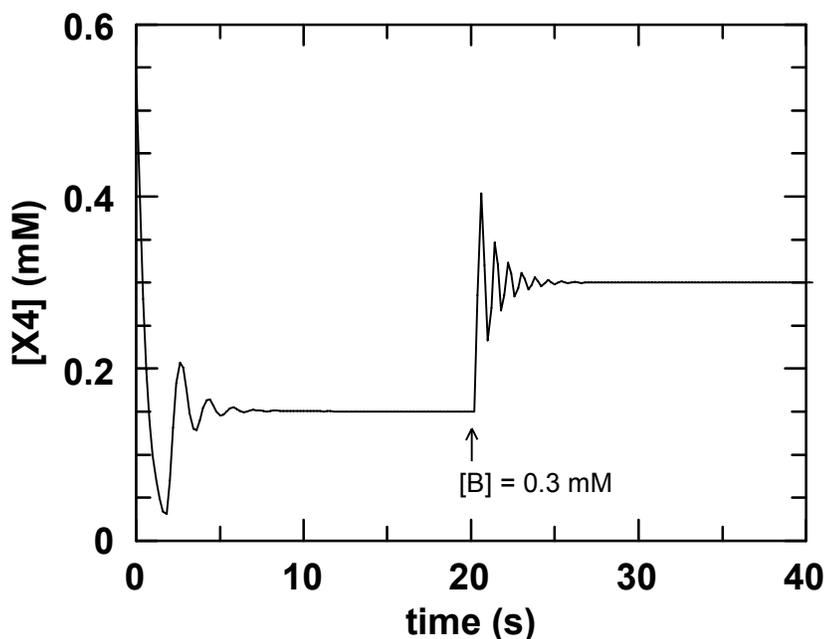
This result could hardly be deduced from a simple causal analysis. In this simulation, Y2 begins at some high concentration and at the start of the run is switched to a value of 0.15 millimoles (mM). The scaling in the simulation is such that Y1 has the same concentration as X4, although signal-molecule concentrations would normally be only a small fraction of the concentrations of primary metabolic substances. We can see that the concentration of X4 (and Y1) first drops, then rises, then quickly settles down at a value close to 0.15 mM. The numerical record of the simulation shows that the final value is *exactly* 0.15 mM, to better than one part in a thousand.

Then, at a simulated time 20 seconds later, Y2 is switched suddenly to a concentration of 0.3 mM. After a few rapid oscillations, the concentration of X4 comes to (exactly) 0.3 mM. So, ignoring the rapid oscillations (they can be eliminated), what can we say that this biochemical system *does*?

Note that when Y2 is set to 0.15 mM, X4 is rapidly brought to a concentration of 0.15 mM, and when Y2 is set to 0.3 mM, X4 is brought quickly to that new concentration. It is reasonable to assume that there is some range over which varying the concentration of Y2, not too rapidly, will make X4 vary in precisely the same way (a control engineer might recognize this as a servomechanism). As a bonus, this system also protects X4 from disturbances of various kinds. Altering the concentration of X1 over a wide range has no significant steady-state effect on X4, even though X4 is one of the products of X1. And changing $k_4$, which represents a drain on X4, also has almost no steady-state effect on X4 over a significant range of $k_4$. Thus a negative feedback control system can be used to set a molecular concentration involved in a main metabolic path to a specific value and keep it there in a varying environment. Clearly, to recognize these basic phenomena of negative feedback control is to open the door to some very new interpretations of what we observe.



Fig. 2. Simulation of system in Fig. 1.

## ORGAN-LEVEL CONTROL SYSTEMS

Walter B. Cannon, early in the 20th Century, invented the term "homeostasis," a term that has been known widely for almost three quarters of a century. Not so well known is the term "rheostasis," introduced by Nicholas Mrosovsky[4]. Both homeostasis and rheostasis are evidence of biochemical control systems, but now at the level of organ systems rather than detailed biochemical reaction dynamics.

One well-known homeostatic system regulates the concentration of thyroxin circulating in the bloodstream. Thyroxin comes from the thyroid gland, which is stimulated to produce it by thyroid-stimulating hormone or TSH. The higher the concentration of TSH in the bloodstream, the greater the rate at which the thyroid gland secretes thyroxin into the bloodstream.

TSH is secreted by the pituitary gland. There are two major influences on the production rate: stimulation by messenger molecules (TRH, or TSH-Releasing Hormone) produced by neural signals reaching the neural part of the pituitary, and suppression by circulating thyroxin molecules reaching the pituitary through the bloodstream (negative feedback). The homeostatic aspect of this system comes from the negative feedback loop: if something such as injecting thyroid extract tends to raise the level of circulating thyroxin, the increasing thyroxin reduces the production of TSH by the pituitary, lowering the TSH concentration and reducing the output of the thyroid gland. A decrease in thyroxin concentration has the opposite effect: more TSH and more thyroid output. The overall effect is to stabilize the level of thyroxin in the bloodstream: hence the "stasis" in "homeostasis."

Essentially every organ system in the body works this way, with various parts of the pituitary gland participating in those comprising the endocrine system. A product of an organ feeds back ultimately to inhibit its own production, with the result that its concentration is stabilized, or as physiologists say "defended," against various kinds of disturbances.

Mrosovsky's book contains a long list of homeostatic systems, but its main point is something else: the set point or defended level or reference level of the stabilized variable is, under many conditions, *itself variable*. The idea of homeostasis applies only over the short term; on a longer time scale, we find that the reference state is quite often, and maybe always, adjustable. Rheostasis, as in rheostat.

Mrosovsky discusses many examples of rheostatic systems, including the thyroxin control system. When an organism is put on a reduced diet, eventually the level of circulating thyroxin **hormones** drops by as much as 50% (Mrosovsky op cit, p. 88). The TSH level still varies within the normal range. Thyroxin concentration continues to be controlled at this lower level, resisting disturbances tending either to increase or decrease it. So evidently the reference level in the pituitary (set by the concentration of TRH) has been reduced, which means, presumably, that the neural signals determining it have been set to lower values by centers in the hypothalamus where those signals arise.

What would cause the reference level of a homeostatic system to vary? Mrosovsky offers a hint: some higher-order process which uses the whole homeostatic system as its effector. Since the homeostatic control loop is already controlling a variable of interest, a higher system that needs to manipulate the same variable would first have to disable the homeostatic controller if it were to act directly on that variable. Rodney Brooks' "subsumption" architecture[5] works this way. But the higher system can easily alter the variable simply by altering the reference signal that tells the homeostatic system the level at which to hold its controlled variable. We can see the beginnings of a *hierarchical control* architecture, in which one system acts by varying the reference signals of several lower systems. And of course they, in turn, can act the same way to use still lower-level systems such as the allosteric-enzyme biochemical control system we saw above. It is also possible for higher systems to monitor the *quality of control* achieved by lower systems, and to act by varying their parameters as well as their reference signals: adaptive control.

There are phenomena like these throughout the body's organ systems. But we move on now to still higher levels, quite possibly skipping some levels, in what is beginning to make sense as a very extensive hierarchy of control systems.

## SPINAL-LEVEL CONTROL SYSTEMS

John Dewey, over 100 years ago, recognized that there is something peculiar about the so-called "spinal reflexes." He realized that the stimuli which seem to elicit them act on sensory nerves which also, almost instantly and indeed while the stimuli are still acting on them, are affected by the motor responses they are producing. To Dewey, it was obvious that the simple concept of stimuli causing reflexive responses was too simple. Instead, he said[6], we have to think of the reflex arc as a complete circle (or as control engineers later would come to say, a feedback loop).

Consider the lowliest of all spinal reflexes, the Golgi tendon reflex[7]. Any force generated by muscle fibers due to signals from the spinal motor neurons excites Golgi tendon organs, which generate sensory signals. Those signals return to the spinal cord where, uniformly, they *inhibit* the same spinal motor neurons that are generating the signals that are causing the muscle to generate a force. When a steady muscle tension is being maintained, there is a continuing feedback signal and a continuing inhibition of the motor neurones. Of course something must also be exciting the motor neurons, to produce any tension to create the negative feedback signals.

Clearly, we have the same situation we have seen at the organ and the biochemical levels. The exciting signals correspond to Y2 in the biochemical control system. The inhibitory feedback signals correspond to Y1, and the muscle tension corresponds to the concentration of X4. The spinal motor neuron, affected both by the excitatory input and by the negative feedback signal, corresponds to the enzyme which is affected positively by Y2 and negatively by Y1, and in turn affects the controlled variable X4. Again, once we know what to look for we find obvious negative feedback control, the same architecture we have seen now at two lower levels.

At the spinal level there are also muscle-length and length-rate-of-change control systems, together making up the stretch control system (commonly called the stretch reflex). These systems act by altering the net excitatory signal entering the tendon-force control system, in a quasi-hierarchical manner. They are most useful when a limb is free to move, whereas the tendon system that appears hierarchically below them can regulate applied force when the limb is constrained and the muscle length control systems are ineffective (isometric operation).

## BEHAVIOR-LEVEL CONTROL SYSTEMS

When centers higher in the brain issue commands to the muscles, those commands appear either as alpha-efferent reference signals that set reference levels for applied force, or gamma-efferent reference signals that set reference levels for muscle length or rate of change of length. No command from the brain is simply relayed to the muscles via the spinal motor neurons: the control loops are always there, strongly affecting the net signal going to the muscles. But the brain does not have to disable the spinal control systems when it needs to produce actions. Instead, it *uses* them by adjusting their reference signals. It tells the control systems not how much to contract the muscles, but what tension or what muscle length to sense. This means that any higher systems stand in hierarchical relation to the spinal control systems, using whole spinal control systems as effectors. This is quite clearly rheostasis at the level of spinal reflexes.

This is the fourth level of negative feedback control we have examined: biochemical control, organ-level control, spinal-reflex-level control, and now what we can call behavior-level control. We have reached the higher reflexes, such as the iris reflex, the balance reflex, and others. But what we see goes much farther than that: we see control loops in which the variables being controlled are located outside the nervous system and muscles, or even in the environment. The controlled variables are now sensed in ways that involve, or can involve, complex perceptual interpretations and even consciousness. The means of controlling them consists of the entire musculature and all the motor control systems that operate the body, and what is controlled is now known to the organism simply as the world of experience. We have entered the realm where behavior is a process by which the organism uses its motor systems to control the states of perceived variables of all kinds.[8, 9, 10]

## CONCLUSIONS

Negative feedback control is not a new principle, but as far as the sciences of life are concerned it is an underutilized principle, mentioned by many but fully understood by few. Many people have suspected the existence of some such architecture, but the mainstream has never been willing to give up the causal model, at least not to an extent sufficient to encourage a major commitment of resources to the study of living hierarchically organized negative feedback control systems. Perhaps in this new millennium we will see a return to this basic concept, and finally an understanding of what it can mean to the sciences of life. I suggest that this is the concept that Brooks said was missing.

## REFERENCES

1. Brooks, R. The relationship between matter and life. *Nature* 409, 409-411 (2001).
2. Hayashi, K. and Sakamoto, N. *Dynamic analysis of enzyme systems* (Springer-Verlag, New York, 1986).
3. "GEPASI" by Pedro Mendes of the University of Wales at Aberystwyth. See URL http://www.gepasi.org/.
4. Mrosovsky, N. *Rheostasis: The physiology of change*. (Oxford University Press, New York, 1990)
5. Brooks, R.A. "How To Build Complete Creatures Rather Than Isolated Cognitive Simulators," in *Architectures for Intelligence* K. VanLehn (ed), 225—239. (Erlbaum, Hillsdale, NJ, Fall 1989).
6. Dewey, J. The reflex arc concept. *Psychological Review* 3, 357-370 (1896)
7. Jami, L. Golgi tendon organs in mammalian skeletal muscle: Functional properties and central actions. *Physiological Review* 72, No. 3, 623-665 (1992).
8. Powers, W.T., Clark, R.K. & McFarland, R.L., "A general feedback theory of human behavior: Part I," in Ludwig von Bertalanffy and Anatol Rapoport, eds, *General Systems: Yearbook of the Society for General Systems Research* **5,** 63-73 (Society for General Systems Research, Ann Arbor, 1960)
9. Powers, W.T. *Behavior: the control of perception* (Aldine/Walter de Gruyter, Hawthorn, New York, 1973).
10. Powers, W.T. *Making sense of behavior* (Benchmark, New Canaan CT, 1999).

Further reading:
   Savageau, M.A. *Biochemical systems analysis: a study of function and design in molecular biology* (Addison-Weseley, Reading MA, 1976)