Homeostatic control mechanisms

Jim Waterhouse

Abstract

Although exact values for physiological or biochemical variables show inter-individual variation, it is important for each value to be kept within a narrow range that is compatible with health. Such homeostasis must be achieved in spite of tendencies for changes to be produced by an individual's environment and lifestyle. Tight and rapid control requires the existence of specific mechanisms: negative-feedback loops. These mechanisms form the basis of physiology, and this article deals with what they are, how they work, and the problems that can arise with them, and how these problems are solved in humans. The main function of such loops is to maintain a constant environment for the cells of the body, but they can also produce controlled change in a variable; for example, if a smooth movement (change of muscle length) is to be undertaken.

Keywords Constant environment; controlled change; feedback loops; oscillation; set point

Learning objectives

After reading this article you should be able to:
- define the term 'homeostasis' and explain its importance to human physiology
- describe how many biological changes spontaneously correct
- describe the components of a feedback loop and outline their functions
- illustrate examples of negative- and positive-feedback mechanisms
- describe what is meant by the 'set point' of a feedback loop and the value of being able to change it

Humans, like all other living organisms, are involved in a continuous struggle to survive not only the rigours and vicissitudes of the environment, but also the biochemical and physiological consequences of any activities. There are constant changes in the ambient temperature, wind speed and humidity. The provision of energy for metabolism also changes after eating or fasting, as does the distribution and amount of blood required by different tissues during exercise. Enzyme-mediated processes work only within a restricted range of pH, temperature and ion concentration, and it therefore becomes necessary to preserve the internal environment despite the surroundings. This was recognized about 150 years ago by Claude Bernard, and about 70 years ago Cannon introduced the term 'homeostasis' to describe this principle.

Variation between and within individuals

If any biochemical or physiological variable is measured on several occasions in the same individual, it is found to vary. These changes can arise from differences in the environment and from an individual's recent activities. For these reasons, it is necessary to standardize the conditions before a measurement is taken. For some variables, the previous effects of lifestyle and the environment take a long time to wear off (e.g. the increase in haematocrit that follows training at altitude) while, for others, there is a regular and reproducible daily change in value due to the activities of a 'body clock' (e.g. the blood level of cortisol).

When more than one healthy individual is studied, interindividual variations are found. It then becomes important to determine if a particular individual falls within the normal range associated with health. A normal value is one that falls inside a range that encompasses 95%, or some other proportion, of the population.

Even if homeostatic control processes are operating on the variable being controlled, it continues to show changes, but they are within a narrower range than would exist without such control. For example, whether people are asleep or exercising, core body temperature does not normally move outside a range of 35–40°C even though they might be in an environment that is hotter or much cooler than this range. Homeostatic processes therefore limit the range shown by a variable and control its mean value.

Spontaneous corrections of some changes

Many changes that normally occur in the body tend to correct automatically without the need to invoke special control mechanisms. If systems in equilibrium are disturbed, changes occur spontaneously to minimize the disturbance.

- Consider the delivery of oxygen from the red blood cell to an active tissue, a process that normally shows a dynamic equilibrium between its different stages. Increased oxygen use by the tissue lowers the concentration of oxygen in the tissue, and this causes oxygen to diffuse more quickly down the increased concentration gradient between the red blood cell and the tissue; in this way, more oxygen is delivered to the active tissue. Such changes reduce the concentration of oxygen inside the red blood cell, so upsetting the original equilibrium that existed between free oxygen, reduced haemoglobin and oxyhaemoglobin in the cell; therefore, more oxyhaemoglobin dissociates, so reducing the fall of free oxygen in the red blood cell. In such a dynamic system, any initial disturbance results in further changes throughout the system that reduce the initial disturbance.
- Enzymes greatly speed up the rate of attainment of equilibrium in biochemical systems. The kinetics of enzyme activity are such that the rate at which the enzyme-mediated reaction occurs increases if the amount of substrate increases.

Jim Waterhouse DPhil DSc was formerly Professor of Biological Rhythms at the Research Institute for Sport and Exercise Sciences, Liverpool John Moores University.

- The rate at which sodium is pumped out across the cell membrane is increased if the concentration of sodium ions in the intracellular fluid increases.
- The rate at which a reaction takes place (in order to regain equilibrium) can be considered as a function of the amount of disequilibrium present at that moment. If blood urea rises above normal then more will be filtered by the kidneys and, because urea moves passively, more will be lost in the urine; the greater the rise of blood urea above normal, the greater the rise of urea excretion in the urine above normal.

These examples illustrate that there are widespread mechanisms by which changes away from an equilibrium position spontaneously reverse without the need for special biological reflexes. However, there would be serious problems if these were the only mechanisms involved. One difficulty arises because there would be no control over what the equilibrium would be. On a hot day, for example, the body would equilibrate at a value that was too high for optimal enzyme activity, and on a cold day at too low a value. Also, after a meal, the whole body would equilibrate with glucose levels that were too high, and during fasting the equilibrium value would fall progressively. There is therefore a need for mechanisms that can control a variable at a value determined by physiological and biochemical considerations rather than one influenced by the surrounding environment. The second difficulty with such mechanisms is that, as equilibrium is approached, the response becomes progressively less; small deviations from the equilibrium value can be corrected only slowly. Therefore, although the above mechanisms are widespread and important, there must also be mechanisms for controlling the organism independently of the environment and for correcting any errors more quickly. These control mechanisms involve feedback loops.

Feedback loops

The principle of feedback control is widely used in engineering (from which much of the terminology is derived), and is often illustrated by reference to the temperature of a thermostatically controlled water bath. In this system, the temperature of the water in the bath (the controlled variable) is continuously monitored by a thermometer (the sensor), which sends a signal to a comparator (Figure 1). As its name suggests, the comparator compares the signal from the sensor with a pre-set value (the set point). If the signal from the sensor falls below this pre-set value (the temperature of the water has fallen), the comparator sends a signal to switch on a heater (the effector or controller). As a result, the water temperature rises and the original disturbance is corrected. The sequence of events is called a feedback loop, because it is initiated by a change in a variable (water temperature) and results in a change to the same variable. There are negative- and positive-feedback loops.

Negative-feedback loops are the backbone of homeostasis. In negative feedback (Figure 1), the initial disturbance initiates changes that negate the original effect. If the initial disturbance in Figure 1 had been a rise in water temperature, a negative-feedback loop would have initiated a cooling process. The action of a negative-feedback loop is to stabilize the controlled variable.

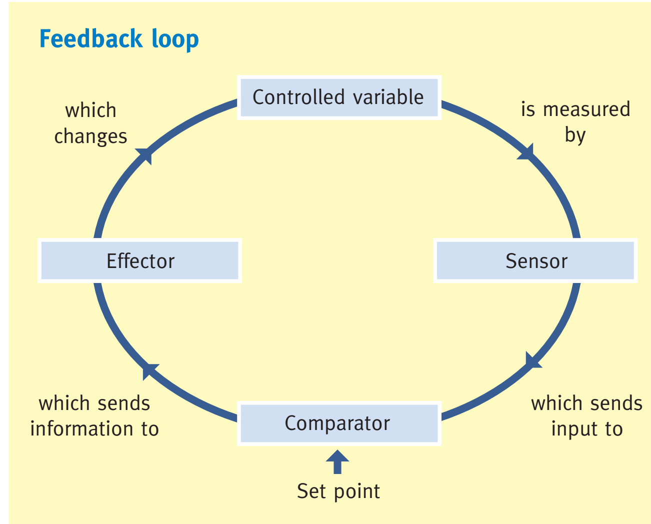


Figure 1

Positive feedback produces changes in the same direction as the initial disturbance. This accentuation of the disturbance acts as a further stimulus to change brought about by the loop. The end result is an increasing change that is limited only by the maximum values that can be achieved by the components of the loop, or by the initiation of some opposing mechanism once a threshold has been reached. There are few examples of positive feedback in human physiology, but they are important.

- Oxytocin release is promoted by the pressure of the fetus on the cervix; this hormone promotes uterine contraction, thus increasing the stimulus for oxytocin secretion. The endpoint is the expulsion of the fetus from the uterus.
- Many gut enzymes are secreted in an inactive form. A small amount of breakdown of this inactive enzyme in the gut produces an active form which, among other actions, promotes breakdown of more of the inactive form into the active one; this process is known as autocatalysis.
- At the onset of inspiration, a group of neurons begin to fire; this promotes activity of the inspiratory muscles and inhibits the firing of expiratory neurons. This activity of the inspiratory neurons also promotes their own firing rate and that of other inspiratory neurons. The result is that, once started, inspiration rapidly takes over from expiration (the same system exists for expiratory neurons and the process of expiration).
- During the rising phase of the nerve action potential, sodium influx across the nerve membrane increases and this causes a depolarization of the cell membrane. In turn, this causes the voltage-gated sodium channels to begin to open, thus increasing sodium influx and a further opening of the gates. This ‘explosive’ process causes the sodium channels to open quickly — the permeability of the cell membrane to sodium increases more than 1000-fold in under 1 millisecond.

Types of feedback mechanism

In the example of the water bath, the heater was either on or off, with no intermediate state; such systems are described as discontinuous. They are less common in physiological control systems where many processes show continuous or tonic activity.

The above examples of positive feedback can be considered as discontinuous systems, with an action potential, expulsion of the fetus, respiratory neuron activity and enzyme activity not being required all the time, but being brought into action when necessary. When called into action, they must work quickly and effectively (this is where positive feedback is so valuable) and the term ‘all-or-none’, which is applied to the action potential, could also be applied to the other situations.

Discontinous systems in biology can also show negative feedback, but they seem to be more of an ‘emergency’ system.

- The clasp-knife reflex is initiated by Golgi tendon organs when the tension in muscles and tendons threatens to become damagingly high. The receptors are comparatively ineffective in normal movements, but when they do become active they inhibit the muscle contraction, so relieving the tension and acting as part of a negative feedback loop.
- Sweating is used when the heat load on the body is too high for it to be controlled by changes in cutaneous vascular tone alone.

Set point

The set point is the value towards which the feedback loop regulates the controlled variable. The set point in a discontinuous negative-feedback system is the threshold at which the system is turned on (e.g. the tension when the effect of the Golgi tendon organs becomes dominant, or the core temperature when sweating starts).

The concept of a set point cannot be applied so easily to a positive-feedback system. However, positive-feedback mechanisms are initiated when, for example, sodium influx across the nerve membrane is sufficient to cause the voltage-gated sodium channels to begin to open or when the neurophysiological state that causes inspiratory (or expiratory) neurons to begin to fire is reached.

The set point of a continuous negative-feedback system is an important concept, and requires the comparator and set point elements of Figure 1 to be replaced by an integrator. This component of the loop is normally found in the central nervous system (CNS) and can be an assemblage of structures. Its role is to integrate the inputs from the various sensors and then to send a signal to the effector systems. Figure 2 shows how the set point can be determined.

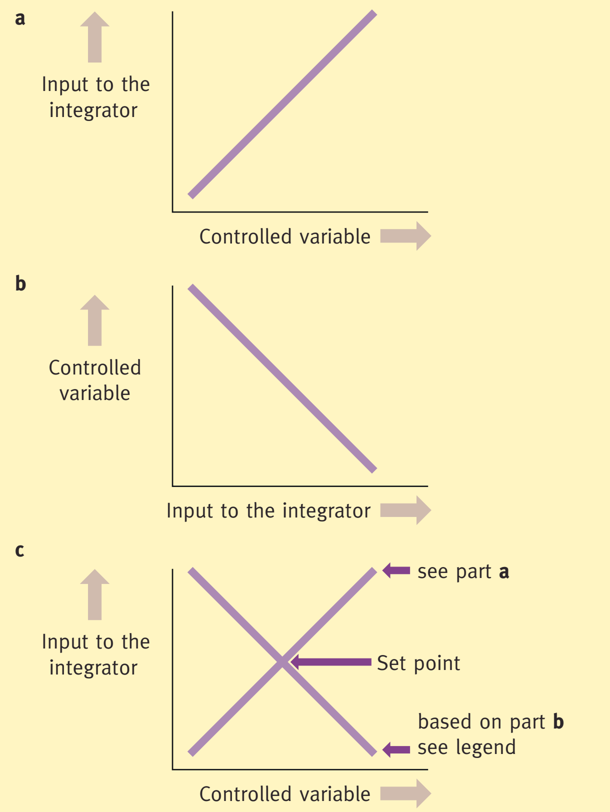
The set point exists when the two parts of the feedback loop balance each other. If the value of the controlled variable differs from this set point, then the two arms of the loop are not in equilibrium and the feedback process will initiate changes that continue until this equilibrium is re-established.

The crossover point (Figure 2) can be considered as indicating the set point of the system and it can be altered by changes to the gradient or intercept of either line; biologically, this means by changes in the properties of the sensor, the central integrator and/or the effector mechanisms. It can become (see later) the mechanism by which alterations to the controlled variable are brought about both in health and in physiological disorder.

Oscillation

Even though a variable is controlled by a negative-feedback mechanism, its value is not constant, and it oscillates (although this does not mean that it is caused by an oscillator) within a band of values. It is desirable that these deviations from the set point, often referred to as ‘errors’, should be as small as possible. There are three main causes of oscillation in negative-feedback systems.

Determining the set point



(**a**) Part of the feedback loop shows a direct relationship between the value of the controlled variable and the input to the integrator. (**b**) Because it is a negative feedback loop, the other part of the loop (which starts with the input to the integrator and ends with the changes to the controlled variable produced by the effector systems) shows an inverse relationship. (**c**) Since the two parts of the loop work are connected and have the same elements (controlled variable and input to the integrator), this can be represented by combining the graphs. The graph of **b** has been reflected and rotated clockwise by 90° before superimposition on **a**. The crossover point indicates the set point of the loop

Figure 2

The ‘store’ of the variable being controlled: in the control of body water, if the store of water is large, then any inbalance between intake and output will produce only a small proportional change in the overall store, its value will move away from the set point comparatively slowly, and there will be more time to correct the error. By contrast, with a smaller store, the proportional change and the consequent rate of movement of the controlled variable away from the set point will be more rapid, with less time available to correct the error before it becomes life-threatening. This would pose a problem to babies, for example.

Sensitivity of the components of the feedback loop: the feedback loop relies for its operation on a sensor that can measure any change in the controlled variable, and integrator and effector mechanisms that can respond readily. If the sensitivity of any component of the loop were too low, then the error would have to become large before compensatory changes were initiated. Problems can arise in babies (in whom many mechanisms are not fully developed), and in old people (in whom mechanisms are deteriorating).

To increase the sensitivity of the feedback loop to error might seem a good solution, particularly since the initial deviation of a controlled variable from the set point is small; with an increased sensitivity, it might be argued, small errors would be dealt with more rapidly. To an extent this is correct, but too high a sensitivity of a feedback loop can also cause oscillation. The oscillation arises because a small error initiates too large a correction response, and so the error becomes one in the opposite direction; this, in turn, initiates a correction in the opposite direction, which also causes the controlled variable to overshoot, and so on.

Delay in the response to an error can be illustrated by attempts to take a shower. If the water is too cold, we turn the dial controlling the water temperature towards hot. Because there is a delay between doing this and the warmer water coming out of the shower, the dial tends to be turned too far towards hot; eventually the water becomes too hot. There is then a tendency to turn the dial too far towards cold, and, after a delay, the water is once again too cold and the cycle begins again. Some computer games require movements of a joystick to track a moving target; if there is a delay between moving the joystick and producing a movement, it can become difficult to perform the task successfully, because there is a tendency to over-correct any error and initiate oscillation around the target.

Neural feedback loops, such as the stretch reflex in muscles, tend to suffer less from this cause of oscillation, because the delay in the feedback loop is small. In hormonal feedback loops, oscillation due to the slower timing of a response is more likely. This is not only due to the time it takes for the secretion by a gland to build up to a sufficient concentration in the whole of the extracellular fluid, but also because, after a reduction of hormone secretion, it takes time for the hormone that has been released to be removed by metabolism.

Minimizing problems

Many of the above problems are minimized in healthy adults. Some of the mechanisms involved are described below.

Multiple and opposing systems: a variable is seldom controlled by a single feedback loop; instead, there are generally several loops acting in parallel, sometimes in opposite directions and with different time courses. The short-term control of blood pressure is through the interaction of loops involving the sympathetic and parasympathetic nervous systems. They tend to have opposite effects on blood pressure, and also act at several sites (heart, arterioles and venules) and by more than one mechanism (inotropic and chronotropic effects on the heart). The long-term control of blood pressure involves several hormones that modify salt and water balance. Several hormones are involved in the control of metabolism; between them they promote the metabolism or storage of glucose, or its production from carbohydrate stores or non-carbohydrate sources, and the storage of fatty acid in adipose tissue or its mobilization from this tissue in preparation for metabolism.

Dynamic sensitivity of sensors: the problem that an error in a feedback loop is initially small can be overcome if the sensor shows dynamic as well as static sensitivity. Dynamic sensitivity is a sensitivity to the rate of change of the variable, and it differs from static sensitivity, which measures the actual value of the variable. When an error first arises, even though the absolute size is small, the rate at which it is changing can be large. Therefore, a sensor with dynamic sensitivity can send a large signal to the integrator as soon as an error appears. Many sensors show dynamic sensitivity, including the baroreceptors and the stretch receptors in muscles.

Disturbance detectors: so far we have considered sensors that are an integral part of the feedback loop in so far as they are responsible for directly measuring the controlled variable; they are sometimes called 'error detectors'. However, another type of sensor is found called a 'disturbance detector'. Even though such a sensor is not a direct part of the feedback loop (and so would not appear in Figure 1), it sends information to the loop, generally to the integrator. Its effect is to give warning of changes that will, in due course, alter the controlled variable.

- A well-known example is in thermoregulation where the error sensors are in the bloodstream (and measure hypothalamic temperature), but the disturbance detectors are in the skin (the cutaneous thermoreceptors). As a result of inputs to the hypothalamus from the disturbance detectors, sweating starts as soon as an individual is placed in a hot environment, and shivering as soon as he/she is in a cold one.
- Another example is the release of gut hormones that affect metabolism, which occurs soon after eating a meal but before most of the food has been absorbed.

In both these examples, the disturbance detector enables the feedback loop to initiate changes before the disturbance has produced an 'error' in the controlled variable.

- A third example is the sensation of thirst. We feel thirsty because of dehydration, and the raised osmotic pressure of the blood acts on osmoreceptors in the hypothalamus. Several effects are produced, including the sensation of thirst and secretion of antidiuretic hormone (vasopressin), which reduces further water loss by the kidneys. After drinking about the amount of water required to correct the osmotic pressure of the blood, the sensation of thirst is abolished, even though the water has not been absorbed and the osmotic pressure of the blood is still raised. The inhibition to drink comes from volume receptors in the stomach and from other inputs originating from the mouth and oesophagus; these can be considered to be acting collectively as a disturbance detector which indicates that, after absorption of the water, the osmotic pressure will have been corrected.

Using feedback mechanisms to produce a controlled change

The set point or equilibrium value of the controlled variable in a negative-feedback system can be controlled by altering the properties of its components (Figure 2). One case where this is normally done is in the performance of a controlled movement; the power for the movement comes from the α-motor units, but control comes from a modification of discharge of γ-efferents. The γ- efferents act on the muscle spindles to change their dynamic and static sensitivities. As a result, a new equilibrium point (crossover point in Figure 2) is achieved, and this causes the controlled contraction of the muscle to take up a new length. Similarly, the sensitivity of the chemoreceptors in the carotid and aortic bodies might be changed in exercise, so that ventilation can be increased without there having to be prior and marked changes in the partial pressure of carbon dioxide and oxygen in the blood. Also, changes in the sensitivity of the integration component of a feedback loop during sleep, and circadian changes due to the body clock, can both result in changes in the set point of several variables (e.g. blood pressure, core temperature).

Changes in the sensitivity of some sensors can occur. For example, the declining elasticity of blood vessels with ageing, particularly those in the carotid sinus, can result in a decreased distension of this structure caused by the blood ejected at systole. This alters the sensitivity of this component of the feedback loop controlling blood pressure, and might contribute to the increase in blood pressure commonly found with ageing.

Interaction between neural and humoral control mechanisms

Humorally mediated reflexes are slower than, but not inferior to, neurally mediated reflexes; they play different roles. Consider the case of the changes in cardiac output that take place during exercise. Initially, the changes are mediated by neural influences on the heart and the vasculature. However, to maintain a high rate of firing of the sympathetic nervous system would be uneconomical from the viewpoint of the metabolic energy required by the sodium—potassium exchange pumps in the cell membrane of the nerve cells. The secretion of catecholamines by the adrenal medulla, itself brought about by sympathetic discharge, can provide an alternative way of maintaining stimulation of the cardiovascular system; it has the added benefit that, as a blood-borne hormone, it can exert its effects throughout the body. In this way, there can be an increase in blood flow to the vasculature of the skeletal muscles, a decrease in blood flow to the gut and kidneys, and a dilatation of the airways — depending on whether the tissue contains mainly α- or β-adrenoreceptors.

Similarly, at the start of a meal, the secretion of saliva needs to be immediate, and is neurally mediated by the autonomic nervous system. As the food passes through the stomach and into the small intestine, the secretory processes (which now do not need to be initiated rapidly but need to be sustained for longer periods) are influenced more by hormonal activation.

In homeostasis, slightly different roles are played by the neurally and humorally mediated reflexes, as can be illustrated by the control of arterial blood pressure. In the short term, it is controlled by the neurally mediated baroreceptor reflexes acting via both branches of the autonomic nervous system. In the long term, it is regulated by hormonal control of the body's water balance (by vasopressin) and sodium content (atrial natriuretic peptide and the renin—angiotensin—aldosterone axis).

It would be wrong to consider that neurally and humorally controlled systems tend to be separate because of their different time courses; they act in parallel. One example is that of the thyroid gland and thermoregulation. Even though thermoregulation is achieved by a series of neurally mediated feedback loops, they act against a background of basal metabolism controlled by thyroid hormones. In cooler weather, there is an increase in secretions from the thyroid gland and these raise metabolic rate. In this way, it becomes less difficult and and economical for the thermoregulatory mechanisms to maintain body temperature.

This combination of fast-acting, neurally controlled and the slower-acting, hormonally controlled mechanisms, requires some point of interaction so that the total response can be integrated. This interaction can be the result of innervation of an endocrine gland by the sympathetic nervous system, as with the islets of Langerhans and the adrenal medulla. In the case of the effects of stress on the release of oxytocin and vasopressin by the posterior pituitary, it is the fact that this part of the pituitary (the neurohypophysis) is composed of neurons.

However, the interaction between the neural and hormonal systems is most elaborate in hormones that come from the non-neural tissue of the anterior pituitary gland. That there is an interaction is established from the interaction between the thyroid gland and thermoregulation (see above), and from the disruption of the menstrual cycle caused by stress, to give two examples.

There are no neural connections to the anterior pituitary gland, but rather a humoral pathway through which releasing factors can travel from the hypothalamus via a portal system. The control of hormone release that occurs through a feedback loop — between a trophic hormone from the anterior pituitary gland, the endocrine gland and the hormone it secretes, and feedback of this hormone on to the anterior pituitary gland — can be modified by the effects of the releasing factors.

The secretion of these releasing factors is controlled by, among other factors, activity in the CNS. The effect of the releasing factor can be considered to be some form of disturbance detector and/or some mechanism for altering the sensitivity of one component of the feedback loop that exists between the pituitary gland and the target gland; whatever description is applied, there is a neuroendocrine integration that benefits the individual. ◆

