4. PRINCIPLES OF MATHEMATICS AND ENGINEERING SCIENCES

The idea of “proof” is the guiding light of mathematics. No matter how many examples you can give for the reality of your theorem, if you cannot offer a valid proof, then your theorem is merely a conjecture.

- Devlin M. Gualtieri

The engineering sciences (such as statics, dynamics, strength of materials, transport processes, and electricity) can be described as refinements or adaptations of physical fundamentals. Some others, however (such as information theory and control systems), are conceptual in nature and lend themselves well to mathematical manipulation. Hence relevant mathematical attributes have been included here along with engineering sciences not directly related to physics.

Intellectual insight can be obtained from mathematical manipulation, and this insight is directly related to the understanding of interactions of biological units with their physical, chemical, and biological environments (Figure 4.0.1). Furthermore, mathematical constructs (such as statistics) can help to describe biological events and predict outcomes.

![Figure 4.0.1. The biological unit (BU) is affected by interactions with its physical, chemical, and biological environment. Likewise, environmental elements are affected by the BU. Self-adjustment is also a possibility. Mathematics and engineering sciences help to describe and analyze these interactions.](image-url)
Although mathematical models are, at best, idealizations, and suffer from a certain degree of unreality, they can still be used to gain understanding and to yield prediction. Engineering sciences are highly dependent on these mathematical models; their idealizations are necessary for simplified analyses of biological characteristics and responses.

Mathematics is not by itself a science, and so has a contribution to the biological sciences that is different from physics and chemistry. It is through mathematics that concepts and organization of principles can be applied; mathematics provides the tools for this process. Some required concepts are not entirely mathematical, however, so we must also draw upon the engineering sciences such as control theory to help understand biological function. In this section, material relevant to biological systems is presented.

1. **There is an element of randomness in biology.** Whether we are looking at some response (such as movement) of a biological system or making a measurement of a physical attribute (such as weight), biological systems seem never to be completely predictable. There are always differences with time, or space, or between individual organisms. If enough was known about the determinants of the measurement, perhaps much of the apparent randomness could be explained, but there is ultimately, at a small enough scale, an inevitable random element always present. Randomness is manageable if the probabilities of occurrences are known for entire populations. See Section 4.2.

2. **Appropriate responses require control systems, and each of these needs sensors, actuators, processing, and information pathways.** Survival requires that adequate responses are made to environmental stimuli. Information important to the biological unit must be sensed, sent on to some structure that determines the appropriate response, and then action taken. These controls may be simple or elaborate, but the fact that they occur at all in biology is a sign of life. See Sections 4.3, 4.4, and 4.5.

3. **Optimization conserves resources.** Most resources necessary for life are limited to some extent, and cannot be wasted if life is to continue to thrive. Optimization is a term that has come to be overused and misused because many things we call optimized are not optimized at all. To be optimum, a process must have some criterion that is to be minimized or maximized, and then the process must minimize or maximize the criterion. Biological systems often have many of these criteria that should be optimized simultaneously. Priorities must then be established. See Section 4.6.
4. *Information implies order.* Shannon’s information theorem that relates information content to the probability of occurrence has been applied mainly to electronic communications systems. Information, however, is extremely important to biological units because information allows the unit to survive, grow, and reproduce. Thus, Shannon’s theorem is relevant to biology, and to maintaining the integrity of biological systems. See Section 4.7.

4.1 Equality

*It depends on what the meaning of “is” is.*

-William J. Clinton

Basic to mathematical manipulations is the concept of equality. This leads to the algebraic notion of an equation, wherein an equal sign is placed between mathematical expressions. There are actually three types of equations, and it is important to understand the differences between them:

1. *Unequivocal equality.* The equal sign in this case means that, under all circumstances, the mathematical expressions on both sides must always be equal. This is a powerful concept. It states that not only must these mathematical expressions be equal in magnitude at some times, or under certain conditions, but they must be equal under all conditions. If there are limits to the range of validity of the equation, then these are stated explicitly outside the equation, and the equation is considered to be true for any set of circumstances within these limits.

   An example of this type of equation is the well-known \( F = ma \), which expresses a fundamental relationship between the force on an object, the mass of that object, and its acceleration. Another example is \( y = 2x + 1 \), where the value of \( y \) is related to the value of \( x \) for all values of \( x \).

   Not only does an equal sign denote equal magnitudes, but it also indicates equal dimensions and identical units for each of the mathematical expressions appearing in the equation. Thus, if one expression is 8 oranges and another is 8 apples, these two cannot be set equal to each other despite the fact that the numbers are the same.

   Energy and power values for biological systems are often expressed in several different sets of units. Among these are watts, calories, kilocalories, British Thermal Units (BTU), ergs, horsepower, electron-volts, and others. Unless the same units are used on both sides of the equation, and unless the same units are used for each expression on one side or other of the equation, then the equation is not valid.

2. *Conditional equality.* The equal sign in this case indicates that equality is achieved only for certain values of the variables, and is
often used as the means to solve for roots. An example of this type of equation is \( x \cot x = 1 \), which is only true for certain discrete values of \( x \) (\( x = 0, 4.4934, 7.7253, 10.9041, 14.0662, 17.2208 \ldots \) (Carslaw and Jaeger, 1959)).

3. **Replacement.** The equal sign in this case means to replace the current value of the left hand variable with the calculation of the term on the right. This kind of equation is used in computer programming, and an example is \( y = x + 1 \). The value of the variable \( y \) is given as the value 1 added to the value of the variable \( x \). The value of the variable \( x \) was previously calculated.

Equations may express theoretical concepts or they may express empirical information. The division between these two is not very clear when closely scrutinized because most theoretical concepts and so-called fundamental principles were based in their formative stages on direct observation. As time went on, and as it became clear that there were means of expressing these observations so that they could generally predict further observations, then they were elevated to the status of principles. These principles may still be subject to modification as further experimental observations cause the principles to be reevaluated. Such was the case for Newton’s laws of mechanics when it became clear that they did not predict actions occurring at the subatomic level.

Some information must remain empirical and will never become a generally accepted first principle. After all, there can only be a small number of fundamental principles (according to the presiding conceptual framework of science). Empirical information is useful nonetheless, especially in the designs of utilitarian devices to be used with living systems. Designs of artificial kidney machines, bioreactors, automobiles, hospital ventilators, and even light bulbs are based on empirical information that serves to produce better products.

Equations are important in the study of living systems because they are often used to express input-output-storage relationships, such as the balance equation 2.2.1. There are numerous inputs and outputs at the interface between a living entity and its environment. Each of these may be the object of considerable study, and there may be no other means to calculate its value than to employ equations expressing all other inputs, all other outputs, and quantities stored in the organism. Indirectly, then, the term of interest may be found.

### Dimensional Analysis

Dimensional analysis has been used in engineering to formulate relationships between various dependent variables and the parameters they depend upon. The results have been useful for analysis and design of products and processes. The basis for dimensional analysis rests upon the
**Dimensional Analysis** cont.

fact, stated earlier, that each term in an equation must have the same dimensions. Mazumdar (1989) has given succinct rules for dimensional manipulations:

1. quantities added or subtracted must have the same dimensions.
2. quantities equal to each other must have the same dimensions.
3. any quantity may be multiplied or divided by any other quantity without regard to dimensions. However, the resulting product or quotient must have appropriate dimensions so that the above rules are not violated.
4. the dimensions of an entity are entirely independent of its magnitude.
5. pure numbers, such as Avagadro’s number, exponents, and the base of natural logarithms (e), have no dimensions.

If all the parameters relevant to a problem are known, then dimensional analysis can be used to form a complete set of dimensionless numbers to formulate a general relationship among output and input parameters. This can be done by expressing the parameters in terms of their basic dimensions and then determining the required form of dimensionless groups.

For example, Mass (M), Length (L), and time (T) are the three basic mechanical dimensions. If we wished to consider the case of a simple pendulum, the relevant variables and their dimensions are:

\[
\begin{align*}
\text{r} &= \text{pendulum length (L)} \\
\text{m} &= \text{mass (M)} \\
θ &= \text{initial angle of displacement from the vertical} \\
&(\text{dimensionless}) \\
\text{t} &= \text{period of oscillation (T)} \\
\text{g} &= \text{acceleration due to gravity (L/T}^2) \\
\end{align*}
\]

Any product of these variables must be of the form:

\[m^a \cdot r^b \cdot θ^c \cdot t^d \cdot g^e\]

where \(a\) through \(e\) are exponents. Dimensions of this group are

\[[\text{M}]^a [\text{L}]^b [\text{]}^c [\text{T}]^d [\text{L/T}^2]^e\]

In order for the product of the above form to be dimensionless,
Dimensional Analysis cont.

\[
\begin{align*}
a &= 0 & \text{from Mass terms} \\
b + e &= 0 & \text{from Length terms} \\
d - 2e &= 0 & \text{from Time terms}
\end{align*}
\]

We have three equations and five unknowns, so we can make two arbitrary choices. If we set \( e = 0 \) and \( c = 1 \), then

\[
a = b = d = 0
\]

giving one dimensionless group equal to \( \theta \).

A second independent dimensionless product is obtained when \( e = 1 \) and \( c = 0 \), yielding

\[
\begin{align*}
a &= 0 \\
b &= -1 \\
d &= 2
\end{align*}
\]

the second dimensionless group is \( g \frac{t^2}{r} \).

These dimensionless groups are called pi terms, and the number of independent pi terms is equal to the number of variables involved minus the number of dimensions in which those variables may be measured (Murphy, 1950). In this example there were five variables and three dimensions, giving two pi terms. Because the formulation of pi terms in the example above was based on arbitrary choices, the pi terms are not unique. However, they are complete.

The relationship between input and output parameters thus takes the form:

\[
F(\theta, g \frac{t^2}{r}) = 0
\]

where \( F(\ ) \) indicates some functional relationship, usually determined by experiment.

The advantage of this technique is that it can be used to form a general relationship that is more efficient than one obtained from considering each variable by itself. Varying, in this case, any part of \( g \frac{t^2}{r} \) will give the same relationship, whether the variation comes about by changing period of oscillation or pendulum length (it is not likely that we can vary the acceleration of gravity). This technique has been found to be
Example 4.1.1 Flow in the Pulmonary Vein (Johnson, 1999)

The pulmonary vein is a very distensible blood vessel. The vein has a zero-pressure diameter of about 0.5 cm and a length about 15 cm. The mean blood pressure is about 1200 N/m² where blood enters the pulmonary vein, and is about 0 N/m² where the vein empties into the left cardiac atrium. Assume that the distensible tube diameter is given by a simple relationship: 
\[ D = D_0 + D_1p \] (Figure 4.1.1). Determine the tube compliance constant value \( (D_1) \) when the vein carries a volume flow rate of 83 mL/sec.

Figure 4.1.1. An elastic tube distends nonuniformly due to pressure differences within a flowing fluid (Johnson, 1999).
Solution:

The equation for volume flow rate through a distensible tube is (Johnson, 1999):

\[
\dot{V} = \frac{\pi}{640 \mu LD_1} [D^5(L) - D^5(0)]
\]

\[
= \frac{\pi}{640 \mu LD_1} \{[D_0 + D_1p(0)]^5 - [D_0 + D_1p(L)]^5\}
\]

The process of iteration is one that uses an equality to mean replacement of the value on the left hand side of the equation with the value calculated on the right. Thus, one would make an initial guess for \(D_1\) and calculate \(\dot{V}\). When the value for \(\dot{V}\) does not equal the known value, another guess for \(D_1\) is made.

Eventually, the known \(\dot{V}\) value is obtained. At that point, the value used for \(D_1\) is correct. Because the pressure at the distal end of the pulmonary vein is about 0 N/m\(^2\), and the term \(D_0^5\) is small, we can obtain an initial estimate for \(D_1\) by approximating the above equation.

\[
\dot{V} = \frac{\pi}{640 \mu LD_1} [D_1p(0)]^5 = \frac{\pi D_1^4 p^5(0)}{640 \mu L}
\]

Thus

\[
D_1^4 = \frac{(83 \times 10^{-6} \text{ m}^3 / \text{sec})(640)(4.5 \times 10^{-3} \text{ N sec/m}^2)(0.15 \text{ m})}{\pi(1200 \text{ N/m}^2)^5}
\]

\[
= 8.2 \times 10^{-6} \text{ m}^3 / \text{N}
\]

We check this result by calculating \(\dot{V}\) from the equation and comparing against the known value of \(8.3 \times 10^{-5} \text{ m}^3 / \text{sec}\).

\[
\dot{V} = \frac{\pi}{640(4.5 \times 10^{-3} \text{ N sec/m}^2)(0.15 \text{ m})(8.2 \times 10^{-6} \text{ m}^3 / \text{N})}
\]

\[
\times \left[ 0.005 \text{ m} + \left(8.2 \times 10^{-6} \text{ m}^3 / \text{N} \right) \left(1200 \text{ N/m}^2\right) \right]^{-5} - (0.005 \text{ m})^{-5}\]

\[
= 6.4 \times 10^{-4} \text{ m}^3 / \text{sec}
\]
By trial and error, we generate the following values:

<table>
<thead>
<tr>
<th>$D_1 \left(10^{-6} \text{ m}^3 / \text{ N}\right)$</th>
<th>$V \left(10^{-4} \text{ m}^3 / \text{ sec}\right)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.2</td>
<td>6.4</td>
</tr>
<tr>
<td>5</td>
<td>2.3</td>
</tr>
<tr>
<td>3</td>
<td>1.1</td>
</tr>
<tr>
<td>2</td>
<td>0.69</td>
</tr>
<tr>
<td>2.5</td>
<td>0.86</td>
</tr>
<tr>
<td>2.4</td>
<td>0.83</td>
</tr>
</tbody>
</table>

The value for $D_1$ is thus about $2.4 \times 10^{-6} \text{ m}^3/\text{N}$.

**Applications and Predictions**

1. If two things will be said to be equal, then they must have the same numerical values and units.
2. Biological entities will sometimes be said to be equivalent only within certain ranges.
3. Convection heat exchange in the circulatory system will be found by summing all other components of a heat balance equation.
4. Natural frequencies of biological oscillations will be found by finding the roots of certain equations.
5. Pressure balance equations will be used to model the respiratory system.
6. Mass movements into and out of the cell will be found to conform to a mass balance equation.
7. Total energy in a fluid system will be obtained by summing potential and kinetic energy terms.
8. Not all equations are true all the time.
9. When programming a computer, the variable whose value is to be updated must always appear in simple form on the left side of the equal sign.
10. Relationships among several variables can be derived by equating their dimensions.
4.2 Randomness and Probability

*Chance favors the prepared mind.*  

-Louis Pasteur

There are few characteristics more basic to living systems than the element of random variation. There may be two or more measurements of some biological activity that appear to be identical in value, but there are many more that differ. There appears to be a variation that cannot be directly predicted in almost any measurement made on a biological system, and this is the random variation of which we speak. There may come a time when models of biological systems are so complex that they can account for effects of genetic variations, present and past environments, and spontaneous acts of free-will. From these models may come the understanding that all biological activity is deterministic and not stochastic, predictable and not random, but that time is far off. For now, biological variation must be accepted as inevitable.

The scale of measurement can influence the amount of variation seen. There is, for instance, one and only one known planetary ecological system. Within that there are many biomes the number of which may change with time, definition, or physical scale. The numbers of individuals of a certain species within similar biomes will vary either over time or even at some constant point in time. Responses of individuals of a species to environmental factors such as temperature begin to show much more variation, and so on. As the scale becomes finer, the perception of variation becomes greater.

4.2.1 Probability Distributions

*I’m tired of all this nonsense about beauty being only skin-deep. That’s deep enough. What do you want, an adorable pancreas?*  

-Jean Kerr

Mathematicians deal with variation by talking about the probability of occurrence. The probability of occurrence when plotted over the range of possible measurement values forms a probability distribution (Figure 4.2.1). Dealing with variability can be made easier by accepting that variation will occur and then characterizing the probability distribution.

So what is probability? It is the fraction of all the measurements that will occur between two measurement values. It is defined in this way, rather than to say that it is the fraction of all the measurements that assume a particular value, because of the scaling problem illustrated earlier. As the measurement scale becomes finer and finer, the probability of occurrence of any one measurement value falls toward zero. However, the range of measurement values can be made small, and we can then determine the probability that any one measurement will be found within that range.
Figure 4.2.1. Various probability distributions important in biology. The Normal distribution is used for most applications. The t-distribution is used for small sample sizes from a Normal distribution. The Log Normal distribution fits some data better than a Normal distribution. The F distribution is used to check equality of variances, and the $\chi^2$ (chi square) distribution is used to check expected values of data. The curves shown here are for various values of distribution parameters (Barnes, 1994).
There are independent and conditional (or dependent) probabilities of occurrence. Independent probabilities are ones where there is no linkage among several events, and conditional probabilities are just the opposite. Almost all probabilities in living systems are conditional probabilities. Growth rate, for instance, is usually dependent on the abundance of nutrition. There is, however, a range of exhibited growth rates, depending on genetic character and local environment. The probability of the occurrences of different growth rates will depend on many factors, both internal and external to the living system.

There is a tendency for most biological measurements to have a higher probability of occurrence around the average, or mean value. This gives some validity to using mean values of data because most individuals will have values somewhere in the neighborhood of the mean. However, if we wish to be sure that we include most of the population within a range of measurements, then we must consider the spread, or range, of the data. Spread is characterized by the term variance. Data spread over a wider range will have higher variance values.

Rates of diseases, drug use, crime, and other measures have this in common: extremes, whether high or low, are more likely to be found in units with low populations (Figure 4.2.2). This reflects the tendency for larger populations to cluster around the mean. It is important to remember when

![Figure 4.2.2. A map of the counties in the U.S. with the lowest kidney cancer rates (green) and the highest kidney cancer rates (red) demonstrates that both are in rural areas with low populations. This illustrates the fact that there is more rate variation for smaller sample sizes than there is for higher sample size (Wainer, 2007).](image-url)
testing engineering designs that the smaller the sample size tested the more likely it is that the results will not reflect average responses of the entire population (Wainer, 2007).

There are a number of probability distributions that have been found to be important by mathematicians when describing real-world data. These have been idealized by assuming extremely large sample data sets, so that the plotted distributions are smooth curves and the mathematical equations describing them are continuous. For small sample data sets, the probability distributions are far from ideal, do not plot as smooth curves, and cannot be described by continuous mathematical expressions.

By far the most widely assumed probability distribution applicable to biological data is the Normal Probability Distribution, or Gaussian distribution. When plotted, this distribution forms the familiar bell-shaped curve that is symmetrical about the mean. The mathematical expression describing this distribution is:

\[
p(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x-\mu)^2 / 2\sigma^2}
\]  

(4.2.1)

where \(x\) is any particular value of data, \(p(x)\) is the probability of occurrence of this data, \(\mu\) is the mean of the data, and \(\sigma\) is the variance. The term \(e^{-(x-\mu)^2 / 2\sigma^2}\) stands for the base of Naperian logarithms (e) raised to an exponent of \(- (x - \mu)^2 / 2\sigma^2\).

The values for the mean (\(\mu\)) and variance (\(\sigma\)) are not known until they are calculated for a particular data set. Once these two parameters have known values, the probability distribution is determined, and the mathematical expression (equation 4.2.1) can be manipulated to give useful predictive results, at least in an abstract sense.

There are cases where the data cannot be evenly distributed about the mean. Two instances of this are where negative data values are impossible or where there are limited numbers of choices for ordinal values. Other probability distributions have been developed to describe these situations, and they may have more descriptive parameters than just the mean and variance to completely determine the shape and size of the plotted curve.

---

**What is Beauty?**

*The absence of flaw in beauty is itself a flaw.*

-Havelock Ellis

When we see, hear, smell, feel, or taste beauty, we seem to know it. We find some landscapes beautiful, or some sunsets, or the faces of some people, or certain paintings or sculptures. We listen to music that is beautiful,
and certain pieces of music evoke deep emotion within ourselves. Foods are not always distinguished as either beautiful or not, but they can be either attractive or unattractive. Combinations of tastes and textures and aromas mark very attractive foods, and some national haute cuisines emphasize contrasts to enhance attractiveness.

We know what it is when we experience it, but just exactly what is beauty?

Beauty has not been completely defined, and it is often personal. We do know this, however: beauty is formed from some intermediate states between completely predictable and completely random extremes.

Completely predictable landscapes, for instance, are boring, very uninteresting. Completely predictable music is tedious, and, for that reason, many people cannot tolerate music from minimalist composers (Phillip Glass comes to mind as an example). Completely predictable food (tofu, grits, or milk perhaps) has nothing in it to make it interesting, and so it is best used in combination with other foods.

Completely random sights, sounds, tastes, smells, or feels are too chaotic for patterns to be discerned. They cannot be figured out, and so are relegated as noise. There is no information for us in a completely random input, so it is dismissed as of no interest.

In between these extremes, however, are experiences that are mostly predictable, but with enough surprises to keep our attention. Think of some of the best jokes you have heard. You probably had tried to predict the punch line (which would have ruined the joke!), but the ending contained surprise – it was not predictable after all. Think also of some of the worst jokes you have heard: could it be that they didn’t make sense to you, and you dismissed them as too random?

Interpersonal relationships seem to follow the same trend. We seek people who are largely predictable in character but varied enough to be interesting. They are predictable, yet with a random, or surprise, element. The ratio of randomness to predictability has yet to be modeled, and is probably itself a random variable.

The need for balance between order and randomness extends also to scientific studies. If the experimenter controls the experiment completely, then the outcome is known before the experiment is conducted. Conversely, if nothing is controlled, then there is no expected outcome (Shapin, 2004). Scientific discovery depends very much on study controlled enough to be useful yet with enough randomness to allow for unexpected results. This is called serendipity.

Music, food, landscapes, faces, jokes, aromas, textures – the really interesting and attractive ones contain elements of predictability coupled with random surprises. Hence, it seems that we are programmed to deal with some amount of randomness in our experiences. The same is true for other animals (see Section 6.22), and indicates that beauty is, indeed, in the eye of the beholder.
4.2.2 Self-Similar Data

There’s luck in odd numbers  - Samuel Lover
There is divinity in odd numbers. - William Shakespeare

Some biological data do not fit any of these descriptions. These are data that are self-similar at different levels of magnification. Branching patterns of retinal nerve cells, blood vessels in the retina, and airways in the lungs have spatial self-similarity that seem to form repeating patterns from a larger scale to a smaller scale (see also Sections 7.3 - 7.5). The electrical voltage across a cell membrane of a T-lymphocyte and ionic current through cell membrane channels in pancreatic β cells are self-similar in time. These are called fractals, and have the same features over a broad range of sizes or times (Figure 4.2.3). Fractal data have no means, because the mean value changes as the scale of measurement changes; fractal data may have no variances because the variance values do not converge to any particular value as the sensitivity of the measurement changes.

4.2.3 Pseudo-Random Data

The fundamental difference between [engineering and science] is that science acknowledges uncertainty, while engineering avoids it.
-Michael A. Russell

Some biological data (and perhaps much more than we think) is not truly random, but is determined in such a way that it appears to be random. An example of this is the time between beats of chick heart cells, where it can
be determined that the period between beats depends on the previous interbeat interval. Although the pattern of interbeat intervals looks random, it is actually deterministic. Such a system is called chaotic.

Chaos

Chaos is an unfortunately-chosen name for something that is not chaotic, anarchistic, or confused. Chaos is, instead, a term used to mean a response that appears to be random but which, in reality, is deterministic. As an example of this (Liebovitch, 1998), consider the case $x_{n+1} = 3.95 x_n (1 - x_n)$. In this example, the value of $x$ at any sample time ($n + 1$) depends on the value of $x$ at the previous sample time ($n$).
Comparing the time series of sampled data from $x_{n+1} = 3.95 x_n (1 - x_n)$ to the time series of data from a truly random event shows that they both appear to be random. Both sets of data bounce around from higher to lower values, and no pattern is apparent. When, however, the values of $x_{n+1}$ are plotted against the values of $x_n$, the random data show no relationship between the two, but the example data show a parabolic relationship. This plot is called a phase space, and suitably choosing the variables to be plotted reveals the underlying relationship.

Because the value of $x_{n+1}$ depends on $x_n$, the phase space is said to be one-dimensional. A relationship between $x_{n+2}$ and $x_n$ would be called two-dimensional. Chaotic phase space is always low dimensional.

Chaotic systems do not give random outputs, but they have a random appearance. The output is predictable (deterministic) if the underlying relationship is known.

Chaotic outputs are very sensitive to initial conditions. Nearly identical initial values can result in very different final values, because chaotic system outputs depend on previous outputs and inputs.
Scientists and engineers usually consider data from biological systems to be distributed randomly, unless shown otherwise. The use of the Normal Probability Distribution pervades studies of all biological systems at all levels.

4.2.4 Statistics

*What is, is, and it is impossible for the same Thing to be and not to be.*

*John Locke*

*Statistics* is a mathematical tool used to separate random from nonrandom (usually intentional) effects. We have seen in Section 1.3 that the scientific method involves the repeated steps of hypothesis and testing. After the experimental data have been gathered, how can we be sure if the hypothesis has or has not been supported?

Experimental data drawn as a sample from an underlying population is used with the process of *induction* (see Sections 1.3 and 1.7) to determine if the hypothesis is true for the entire population. Because induction involves some elements of guesswork, statistical tests are used as a guide to make the best possible guess about the hypothesis. There are four possibilities (Figure 4.2.4):

1. The hypothesis is really true, and it is supported by the data. In this case, the hypothesis is accepted correctly.
2. The hypothesis is really false, and the data supports its falsity. In this case, the hypothesis is rejected correctly.
3. The hypothesis is really false, but the data indicates that it is true. In this case, the hypothesis is accepted in error. This is called a Type I error.

**Chaos cont.**

Biological systems may act in a chaotic manner. The outcome of a human life, for example, depends not only on genetic predisposition, but also on environmental factors such as nutrition, education, and opportunity. So, the outcomes (accomplishments, life styles, number of offspring, etc.) may be dependent not only on initial conditions but also on conditions along the way.

It would not be in an organism’s best interest to be totally at the mercy of environmental conditions, so biological systems are not totally chaotic. Whenever possible, biological units attempt to regulate their responses through active control mechanisms (see Section 4.4). It can thus be said that biological systems bring order out of chaos.
4. The hypothesis is really true, but the data does not support it. In this case, the hypothesis is rejected in error. This is called a Type II error.

Most statistical procedures are concerned with minimizing Type I errors, and the probability of the occurrence of a Type I error is fixed in advance of the conduct of the experiment (within the limitations of certain assumptions made about probability distributions of errors). The level of one Type I error being made for every 20 experiments (p = 0.05) has generally been accepted by the scientific community. This error rate has usually been designated by the symbol $\alpha$.

![Figure 4.2.4](image)

Figure 4.2.4. An illustration relative to Type I and Type II errors. The top frequency distribution is of the original untreated population. The bottom frequency distribution is of the population after application of the treatment. The treatment changed the mean, but not the variance, of the population. If there were no difference between means of untreated and treated populations, and data were obtained from the shaded area in the upper distribution, one would conclude that the treatment was effective when it really was not (Type I). Alternatively, if there were a real difference between treatment means, but the data were taken from the shaded portion of the lower distribution, then the conclusion that the treatment was ineffective would be false (Type II).

Type II errors are more difficult to predict. The concept of a Type II error is important to determine the sample size of an experiment to detect a difference of stated magnitude.

Statisticians talk about unbiasedness and robustness of their methods. An unbiased procedure is supposed to be a fair and honest estimate of the
parameter, whereas a robust procedure can be applied in nearly all cases without difficulty. Often, both unbiasedness and robustness depend on the assumptions underlying the development of the procedure.

Statistical tests depend upon mathematical models of expected results. Models related to the effect on the mean (average) of the population by the treatment are given in general by:

\[
\text{sample mean} = \text{population mean} + \text{treatment effect} + \text{error} \quad (4.2.2)
\]

The statistical test to accept or reject the hypothesis (treatment has an effect) is based on a test of the sample mean.

There is always error incorporated in every measurement. The error term in statistical tests is usually considered to be random, and the underlying population error is usually assumed to be a Normal distribution. If either of these conditions is violated, then modifications must be made in standard procedures. Statistics is mature enough that many specialized tests have been developed.

The Student’s t test is a widely used test of experimental means. For equal numbers of unpaired observations of two treatments, one of which may be no treatment at all, the sample means (\(\bar{x}_1\) and \(\bar{x}_2\)) for the two samples (\(x_{1i}\) and \(x_{2i}\)) are calculated:

\[
\bar{x}_j = \frac{\text{sum of observations}}{\text{number of observations}} = \frac{\sum x_{ji}}{n_j} \quad (4.2.3)
\]

An estimate of the sample variance is also calculated, and the square root of the variance estimate, called the standard deviation (s), is used to calculate a t statistic:

\[
\text{t}_{\text{calc}} = \frac{|\bar{x}_1 - \bar{x}_2|}{s} \quad (4.2.4)
\]

This value is then compared to a table of t values. If the calculated t is greater than the tabled value, the treatment effect is considered to be statistically significant: the hypothesis is accepted with a \(p = 0.05\) chance of being wrong.

A Normal distribution is also characterized by its variance, and other statistical tests (the F test is one of these) may be used to check whether the treatment has changed the population variance from what it was before the treatment was applied. If the F test is used, the calculated value is compared to the tabled value, in a similar manner to the t test above. Based on this comparison, a decision can be made about the acceptability of the hypothesis that the treatment is effective.

The study of statistics is, for the most part, a study of various tests and their calculations. However, the application of statistics depends very
strongly on a philosophical foundation. In order to avoid bias in the results, the appropriate statistical test must be decided upon before the data are seen. Once the data are known, even partly, it is tempting to apply a particular statistical procedure that seems to fit patterns in the data. Although there is no difference in the outcome of the calculation when done this way, the interpretation of the result can be erroneous. Studying the data before a specific statistical test is applied violates the assumption of random error.

Can what you don’t know hurt you? It can in statistics. Imagine the following scenario: there was a test of the effectiveness of two drugs, A and B. There were 2000 people who were scheduled to take the experimental drugs, but test subjects had to be identified as the opportunity arose, so not all 2000 started at the same time. In the first 300 pairs of subjects, drug A proved much more effective than drug B in 275 of the pairs. Drug B, even proved somewhat harmful in 100 of the test subjects. Because the results were so overwhelmingly in favor of drug A, the test was cancelled, drug A was declared to be superior, and drug B was labeled as harmful.

However, if the test were continued, it might have been found that drug B would have been found to be more effective than drug A in 600 of the next 700 subject pairs. Not only that, but drug A may have been harmful for 200 subjects. Incorrect conclusions were drawn because there is a small, but finite probability that random events can form patterns some of the time. The philosophical assumptions of basic statistics were violated.

Sometimes statistical methods are used to determine overall treatment effects without accounting for individual or group variations. This is often the case with clinical trials of new drugs, where the results of interest are the benefits and risks on the average population. However, drugs are prescribed not for populations, but for individuals who may or may not react as the population does. Groups of individuals can be analyzed for different expected risk and benefit outcomes in the absence of the drug, and these different groups may give entirely different outcomes when given the drug. Most, if not all drugs (or surgical procedures, or medical devices, or treatments, etc.) convey both risks and benefits. If the overall result from a clinical drug trial shows no benefit from use of the drug, this may be because all participants in the trial showed no effects, but the no-benefit average results could also derive from one subgroup benefiting very positively and another subgroup being harmed significantly. How the drug is administered as part of a treatment program can differ greatly as a result of this difference (Figure 4.2.5). Analysis of clinical trial results by risk stratification is very rarely, if ever, done (Kent and Hayward, 2007). However, there is need to be aware of this procedure in many experiments in biology.

Statistical procedures can also be used to design experiments to be most useful with the least cost. There are classical statistical designs to be used for field experiments, for multiple hypotheses to be tested simultaneously, and for control of covariates.

Covariates occur quite frequently, when levels of a second or third variable depend on the magnitudes of a first variable. For instance, growth
rate varies directly, although not exclusively, with nutrition level. Other variables also affect growth rate, disease and genetics among them. Statistical procedures can help separate various effects.

Figure 4.2.5. Here are expected outcomes for two patients at risk for a stroke. If no surgery is performed, patient A has a much higher risk than patient B of developing a stroke. If surgery is performed, the risk of a stroke as a result of the surgery is much higher in patient B than in patient A. Patient A benefits from the surgery, but patient B is exposed to higher risk. If these two patients are not distinguished, (avg), then there is no indication that surgery should be performed on either.

It is often also true that there is a training or a time-dependent effect that cannot be avoided. Just exposing a human, animal, or plant to repetitive experimental conditions results in adjustments that improve performance. This would be called a training effect, and training effects are common in athletic procedures, in learning experiments, and in skill acquisition. Time-dependent effects are always present with living things because they are constantly adjusting, growing, and maturing. Often, time-dependent effects can be ignored if the duration of the experiment is relatively short or if it is long enough to cover the entire life span of the object of the experiment. In between, however, time-dependent effects can appear to reinforce or to counteract effects caused by experimental treatments. One popular experimental design, the Latin Square, attempts to apportion time or training effects equally to all treatment levels (Figure 4.2.6).
Figure 4.2.6. One possible Latin Square arrangement. Entries in the interior cells of the Table are experimental session numbers, given so that each treatment appears the same number of times in each row and column. This arrangement balances training and time-dependent effects as long as all subjects can be considered the same.

Statistical ideas are often used to derive empirical equations from sets of data, and these equations form the basis for mathematical models used by biological engineers (see Section 1.4). The method of least squares is a very popular procedure wherein a curve can be drawn through a set of data in an attempt to extract the essential non-random variation of data responding to various levels of an input variant (Figure 4.2.7). Although a very popular technique, the method of least squares is only valid for data that is linear and has a constant variance. For nonlinear relationships or non-constant variances, modifications of the technique are available, but they, too, have limitations.
Figure 4.2.7. These are four graphs of four data sets with identical statistics, including the best-fit linear line through the data (Johnson, 1991). Only for the first (upper left) data set is the line appropriate. For the second data set a parabolic line would be correct, for the third set the outlier should be ignored, and for the fourth set there is no correct line. These graphs illustrate the point that data should always be seen before statistical procedures are blindly applied.

**Linear Least Squares Method**  
*(Johnson, 1991)*

If we were to locate a line by eye through a data set $y_i$, we would probably try to balance the distance of the line from the data points. We would take distances on the upper side of the line and balance them with distances on the lower side. If we were really good, we would take all data points into consideration, and maybe even decide that a longer distance from the line to that data point way up there is balanced by the many
**Linear Least Squares Method** cont.

shorter distances to the cluster of data points on the lower side of the line.

The least squares method is a mathematical way of doing what was just described. Vertical distances of the line to the data point \((y_i - \hat{y}_i)\) are used. (Here, \(y_i\) represents the \(i^{th}\) data value, and \(\hat{y}_i\) represents the estimated value of \(y\) that lies on the best fit line.) But because distances may be either positive or negative, and so may cancel, the distances are squared, \((y_i - \hat{y}_i)^2\), to make the numbers all positive. Next, all data points are used: \(\sum (y_i - \hat{y}_i)^2\), where \(\sum\) denotes the sum of the squared differences inside the parentheses. To find the parameter value to minimize this quantity, its derivative (see Section 4.3.1) is set to zero:

\[
\frac{d}{da_j} \sum (y_i - \hat{y}_i)^2 = 0
\]

For instance, if \(\hat{y} = a_0 + a_1 x + a_2 x^2\), and we need to find values of \(a_0\), \(a_1\), and \(a_2\) to produce the best-fit line, then

\[
\frac{d}{da_0} \sum (y_i - \hat{y}_i)^2 = \frac{d}{da_1} \sum (y_i - \hat{y}_i)^2 = \sum \frac{d}{da_2} (y_i - \hat{y}_i)^2 = 0
\]

\[
0 = \sum \frac{d}{da_0} (y_i - a_0 - a_1 x_i - a_2 x_i^2)^2
\]

\[
0 = -2 \sum (y_i - a_0 - a_1 x_i - a_2 x_i^2)
\]

\[
0 = \sum y_i - \sum a_0 - \sum a_1 x_i - \sum a_2 x_i^2
\]

\[
0 = \sum y_i - N a_0 - a_1 \sum x_i - a_2 \sum x_i^2
\]

where \(N = \) number of data points. Also

\[
\frac{d}{da_1} \sum (y_i - \hat{y}_i)^2 = 0
\]

\[
0 = -2 \sum x_i (y_i - a_0 - a_1 x_i - a_2 x_i^2)
\]
**Linear Least Squares Method** cont.

\[ 0 = \sum x_i y_i - a_0 \sum x_i + a_1 \sum x_i^2 - a_2 \sum x_i^3 \]

And

\[ \frac{d}{d a_2} \sum (y_i^2 - \hat{y}_i^2) = 0 \]

\[ 0 = \sum x_i^2 y_i - a_0 \sum x_i^2 - a_1 \sum x_i^3 + a_2 \sum x_i^4 \]

Thus, the set of least squares equations to determine \( a_0, a_1, \) and \( a_2 \) values is:

\[ \sum y_i = N a_0 + a_1 \sum x_i + a_2 \sum x_i^2 \]

\[ \sum y_i x_i = a_0 \sum x_i + a_1 \sum x_i^2 + a_2 \sum x_i^3 \]

\[ \sum y_i x_i^2 = a_0 \sum x_i^2 + a_1 \sum x_i^3 + a_2 \sum x_i^4 \]

These can either be solved simultaneously, or more likely, be solved using matrix methods. By noting the patterns that occur in these equations, they can be written without resorting to differentiation.

Cramer’s rule states that the value of a variable in a set of simultaneous equations can be determined by evaluating the value of the coefficient matrix with the column corresponding to the variable of interest, replaced by the last column of the augmented matrix (Sokolnikoff and Redheffer, 1958). For three or fewer variables, this process is particularly easy. The three variables in the set of equations above are \( a_0, a_1, \) and \( a_2 \). In explicit form, the determinant of the coefficient matrix becomes:

\[ D = N \left[ \sum x_i^2 \sum x_i^4 - \left( \sum x_i^3 \right)^2 \right] \]

\[ - \sum x_i \left[ \sum x_i \sum x_i^4 - \sum x_i^3 \sum x_i^2 \right] \]

\[ + \sum x_i^2 \left[ \sum x_i \sum x_i^3 - \left( \sum x_i^3 \right) \right] \]

For \( a_0 \),

\[ N_0 = \sum y_i \left[ \sum x_i^2 \sum x_i^4 - \left( \sum x_i^3 \right)^2 \right] \]
Example 4.2.1 The Meaning of the Mean

When looking at the Normal, or Gaussian, distribution, it is the central hump that beckons for the most attention. It is there that the largest number of individuals is found, and these individuals reprise the group that is considered average, or normal, or usual. Humans that have physical or mental
characteristics clustered around the mean are usually easily accepted by others and share many of these characteristics with their friends and acquaintances. This is the situation of the common man, of whom Abraham Lincoln said that God must have loved them, for he made so many of them.

It is the tails of the distribution that contain interesting and influential individuals. People with disabilities are at one end of the distribution, and it has only been recently that physical access to public facilities has been guaranteed by the U.S. legal system through the Americans with Disabilities Act (ADA). Medicines have been developed at dosages that must be effective for a vast majority of the population, and that means for people including those at the least-sensitive tail of the population. Many medicines, therefore, are often stronger than needed for the vast majority of people.

**Example 4.2.2  Protecting Against Hyperthermia**

Core body temperatures of 40º C have been considered to be where there are 50% heat casualties in armed combat. In civilian occupations this high rate of heat casualties is not even thinkable, much less tolerable. Thus, government regulations meant to protect workers must use a more stringent standard, one based on protecting a higher percentage of the population. Thus, excessive heat strain has been defined in the U.S. as marked by body core temperature greater than 38.5º C for medically selected and acclimatized personnel or greater than 38º C in unselected, unacclimatized workers.

**Example 4.2.3  Digestibility of Corn Silage in Sheep and Steers (Steele and Torrie, 1969)**

Data for digestibility of corn silage (chopped and preserved corn plants) are given for 7 sheep and 6 steers, as follows. Determine if there is a statistically significant difference between the two types of animals.

<table>
<thead>
<tr>
<th></th>
<th>( x_{1j} ) (sheep)</th>
<th>( x_{2j} ) (steers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_{1j} ) (sheep)</td>
<td>( x_{2j} ) (steers)</td>
<td></td>
</tr>
<tr>
<td>57.8</td>
<td>64.2</td>
<td></td>
</tr>
<tr>
<td>56.2</td>
<td>58.7</td>
<td></td>
</tr>
<tr>
<td>61.9</td>
<td>63.1</td>
<td></td>
</tr>
<tr>
<td>54.4</td>
<td>62.5</td>
<td></td>
</tr>
<tr>
<td>53.6</td>
<td>59.8</td>
<td></td>
</tr>
<tr>
<td>56.4</td>
<td>59.2</td>
<td></td>
</tr>
<tr>
<td>53.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Basic sums:

\[ \sum x_{ij} = 393.5 \quad \sum x_{2j} = 367.5 \]
\[ \sum x_{ij}^2 = 22,174.41 \quad \sum x_{2j}^2 = 22,535.87 \]
\[ \bar{x}_1 = 56.21\% \quad \bar{x}_2 = 61.25\% \]

Calculation of sample variances:

\[ \sum \left( x_{ij} - \bar{x}_1 \right)^2 = \sum x_{ij}^2 - \left( \sum x_{ij} \right)^2 / n_1 \]
\[ = 22,174.41 - 22,120.32 \]
\[ = 54.09 \]
\[ = (n_1 - 1)s_1^2 \]

\[ \sum \left( x_{2j} - \bar{x}_2 \right)^2 = \sum x_{2j}^2 - \left( \sum x_{2j} \right)^2 / n_2 \]
\[ = 22,535.87 - 22,509.37 \]
\[ = 26.50 \]
\[ = (n_2 - 1)s_2^2 \]

Estimating the common variance:

\[ s^2 = \frac{\sum \left( x_{ij} - \bar{x}_1 \right)^2 + \sum \left( x_{2j} - \bar{x}_2 \right)^2}{(n_1 - 1) + (n_2 - 1)} \]
\[ = \frac{54.09 + 26.50}{6 + 5} \]
\[ = 7.33 \]

The common standard deviation for the difference between the two means is:

\[ s_d = \sqrt{s^2 \left( \frac{n_1 + n_2}{n_1n_2} \right)} \]
\[
= \sqrt{7.33 \times \frac{(7 + 6)}{42}}
= \sqrt{2.27}
= 1.51\%
\]

The t statistic is:

\[
t = \frac{(\bar{x}_1 - \bar{x}_2)}{s_d}
= \frac{56.21 - 61.25}{1.51}
= -5.04
= -3.33
\]

The number of degrees of freedom (df) for this example is the total number of data points less the two means estimated:

\[
df = (n_1 - 1) + (n_2 - 1) = 13 - 2 = 11
\]

The calculated value of t is next compared to tabled values, using \( \alpha = 0.05 \) and considering that either value could have been larger than the other, called a two-tailed test.

An excerpt from a Table of t values (Steele and Torre, 1960) is given here:

<table>
<thead>
<tr>
<th>Probability of a larger value of t</th>
<th>0.1</th>
<th>0.05</th>
<th>0.02</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6.314</td>
<td>12.706</td>
<td>31.821</td>
</tr>
<tr>
<td>3</td>
<td>2.353</td>
<td>3.182</td>
<td>4.541</td>
</tr>
<tr>
<td>5</td>
<td>2.015</td>
<td>2.571</td>
<td>3.365</td>
</tr>
<tr>
<td>8</td>
<td>1.860</td>
<td>2.306</td>
<td>2.896</td>
</tr>
<tr>
<td>10</td>
<td>1.812</td>
<td>2.228</td>
<td>2.764</td>
</tr>
<tr>
<td>11</td>
<td>1.796</td>
<td>2.201</td>
<td>2.718</td>
</tr>
<tr>
<td>12</td>
<td>1.782</td>
<td>2.179</td>
<td>2.681</td>
</tr>
</tbody>
</table>
Because the magnitude of the calculated t value (-3.33) is greater than the tabled value (2.201) with α = 0.05 and df = 11, the two means are considered to be statistically different.

**Example 4.2.4 Elastic Properties of Heart Muscle**

In Figure 4.2.8 is shown a diagram of the isometric elastic properties of heart muscle for a 10 kg dog during systole (Johnson, 1991). Develop an equation to describe the pressure-volume relationship of the muscle.

![Figure 4.2.8. Isometric elastic properties of heart muscle for the 10 kg dog during systole (Johnson, 1991).](image)

**Solution:**

A linear least-squares quadratic polynomial equation will be developed. Data were obtained from the curve. For purposes that will be explained later, volume data will be limited to the range of 0 to 29.2 mL. For consistency of notation, pressure will be designated as \( y \) and volume as \( x \).
The curve up to the point where the last data point was taken appears to be an inverted parabola, so a quadratic polynomial will be chosen to represent the curve. From equations in the box on Linear Least Squares,

\[
D = 6[(1167.37)(775019.71) - (28646.26)^2] \\
- 59.7 \left[ (59.7)(775019.71) - (28646.26)(1167.37) \right] \\
+ 1167.37 \left[ (59.7)(28646.26) - (1167.37)^2 \right] \\
= 144512317.53
\]

\[
N_0 = 140 \left[ (1167.37)(775019.71) - (28646.26)^2 \right] \\
- 59.7 \left[ (2132.00)(775019.71) - (28646.26)(45203.00) \right] \\
+ 1167.37 \left[ (2132.00)(28646.26) - (1167.37)(45203.00) \right] \\
= 133483814.55
\]

\[
N_1 = 6 \left[ (2132.00)(775019.71) - (28646.26)(45203.00) \right] \\
- 140 \left[ (59.7)(775019.71) - (28646.26)(1167.37) \right] \\
+ 1167.37 \left[ (59.7)(45203.00) - (2132.00)(1167.37) \right] \\
= 593664070.87
\]

\[
N_2 = 6 \left[ (1167.37)(45203.00) - (28646.26)(2132.00) \right] \\
- 59.7 \left[ (59.7)(45203.00) - (2132.00)(1167.37) \right] \\
+ 140 \left[ (59.7)(28646.26) - (1167.37)^2 \right] \\
= -13715380.27
\]

\[a_0 = \frac{N_0}{D} = 0.92\]

\[a_1 = \frac{N_1}{D} = 4.11\]

\[a_2 = \frac{N_2}{D} = -0.095\]

The least squares equation is thus,

\[y = 0.92 + 4.11x - 0.095x^2\]

To check this equation, pressure data from the curve were compared with pressure data calculated by means of the equation:
The last two data points are outside the range of the original data, and the predicted pressure is very inaccurate. This illustrates one problem with polynomial fits of data: one must be very careful to avoid extrapolation outside the original range.

### Applications and Predictions

1. As the precision of the measurement becomes greater and greater the probability of two measurements being exactly equal will decline.
2. 68% of a normal population will fall within ±1 standard deviation from the mean.
3. As the number of measurements increases, the probability data will look more and more like a Gaussian distribution.
4. Average measurements will apply to only a small number of individuals.
5. No matter how a system is designed to work with living organisms, it will occasionally be inadequate because of random variation.
6. Interactions between macromolecules without the presence of enzymes will depend on random collisions.
7. Increasing the number of observations will decrease the uncertainty of the measurement.
8. Many biological activities will be conditional on the presence or absence of other factors, and so will not be independent.
9. Fetal development is an example of a chaotic system.
10. The larger is the random component of a measurement, the more measurements must be taken.
4.3 Calculus

A mathematician is a machine for turning coffee into theorems.
-Paul Erdos

One of the greatest advances in mathematics came about with the formulation of methods of calculus. Two men working independently, Isaac Newton and Gottfried Wilhelm Leibniz, discovered calculus nearly simultaneously, and the resulting feud over which of the two would be known as the inventor of calculus lasted for many years (Hellman, 1998). Calculus is the branch of mathematics that deals with infinitesimally small changes and the infinite sum of such changes. Thus, calculus deals with things in motion, not only mechanical motion, but also anything that changes with time or space. Because biological systems vary with time and space, calculus has important biological applications.

4.3.1 Derivatives and Differential Equations

Mathematics possesses not only truth, but supreme beauty.
-Bertrand Russell

Derivatives are rates of change of a dependent variable with respect to one or more independent variables. We have already seen derivatives in Section 2.1.2, where capacity was related to the rate of change of voltage with time. Inertia, in Section 2.1.3, was seen to be related to the time rate of change of current.

An equation that contains at least one term with a derivative is called a differential equation. If there is but one independent variable, and the derivative appears only with an exponent of 1.0, then the differential equation is called a linear ordinary differential equation. If there is more than one independent variable, then the differential equation is called a partial differential equation.

4.3.2 First Order Equations

It was ultimately recognized, as Poincaré pointed out, that a complete conspiracy is itself a law of nature, that it is not possible to discover an ether wind by any experiment; that is, there is no way to determine an absolute velocity.
-Richard Feynman

Differential equations may be either first order, or second order, (third order and others are also possible, but less likely), depending on the derivative levels that appear in the equation. For instance, a differential equation commonly appearing in biological systems is the first order constant coefficient linear ordinary differential equation:
\[ x + \tau \frac{dx}{dt} = 0 \]  \hspace{1cm} (4.3.1)

This equation has two terms, a term depending directly on the variable \( x \), and a term involving the rate of change of \( x \) with respect to time \( \frac{dx}{dt} \). Many biological receptors, for instance, have an output that depends on the level of the input and also on the rate of change of the input with time. The variable \( x \) may stand for temperature, glucose level, or light, for example (see Section 6.20.1).

If we try as a solution, \( x = e^{-t/\tau} \), where \( e \) is the base of Naperian logarithms (\( e = 2.718 \ldots \)), then \( \frac{dx}{dt} = -\frac{1}{\tau} e^{-t/\tau} \), and

\[ x + \tau \frac{dx}{dt} = e^{-t/\tau} + \tau \left( -\frac{1}{\tau} \right) e^{-t/\tau} = 0 \]  \hspace{1cm} (4.3.2)

Thus, the equation is satisfied by the trial solution. Situations where the response \( x = e^{-t/\tau} \) are called exponential responses, and are very common in biology at all levels.

### 4.3.3 Exponential Responses

_Thomas Malthus relied on an exponential-growth model to make his famous prediction about human population growth._

_Santiago Schnell_

The rate of unrestricted reproduction of cells depends on the number of cells present. When only a few cells are present, the rate of appearance of new cells is small. As more and more cells are produced, the rate of appearance of new cells increases. This has the effect of producing new cells at a faster and faster rate until such time as reproduction is no longer unrestricted.

Before that happens, however, the growth in the number of cells is considered to be exponential, and can mathematically be expressed as:

\[ \text{number of cells} = (\text{number of cells at time 0}) e^{t/\tau} \]  \hspace{1cm} (4.3.3)

where \( t \) stands for time and \( \tau \) is said to be the time constant. When the value of time is equal to the value of the time constant, the ratio \( t/\tau \) equals 1.0, and \( e^{1.0} = 2.718 \). As the value of \( t \) increases, \( e^{t/\tau} \) becomes greater and greater (Figure 4.3.1).

Most biological responses are not as unbounded as this. Indeed, such unconstrained growth represents the lack of control or balance that
characterizes most biological systems. Exponential processes are still important, but they act to diminish rather than increase the exponential term as time goes on. This can be accomplished mathematically by inverting the term $e^{\frac{t}{\tau}}$, which is equivalent to making the exponent negative, or $e^{-\frac{t}{\tau}}$. As time increases, the magnitude of $e^{-\frac{t}{\tau}}$ becomes smaller and smaller (Figure 4.3.1).

We may consider cell death rather than growth. If the rate of cell death is proportional to the number of cells present, then:

$$\text{number of cells} = (\text{number of cells at time 0}) \times e^{-\frac{t}{\tau}} \quad (4.3.4)$$

Figure 4.3.1. Exponential curves. The upper left curve is an unbounded exponential curve where $t/\tau$ is positive. This curve can be used to represent the unrestricted growth or death of cells. The middle curve is exponentially decreasing, and represents some kind of biological decay. The lowest curve is an exponential response to a step input, and is very commonly seen in biology when conditions change suddenly.
The term $e^{-\frac{t}{\tau}}$ never really assumes a value of 0, except abstractly when time is infinite. Because there cannot be fractional numbers of cells present, when equation (4.3.4) predicts less than one cell, there will be a small probability that a cell survives, but this probability decreases as time goes on. This fact is important in processes to sterilize food or medical products (see Section 6.23).

Biological systems are sometimes asked to respond to sudden changes from one level to another. Such is the case for a sudden change in environmental temperature or a sudden change in peripheral vascular resistance or even a sudden change in interstitial potassium ion concentration. The biological response appears to be exponential in nature, but rather than vary between extremes, as given by equations (4.3.3) and (4.3.4), the response varies exponentially between one level and another. The equation that expresses this response is:

\[
\text{response} = \text{level 1 response} + (\text{level 2 response} - \text{level 1 response}) (1 - e^{-\frac{t}{\tau}}) \tag{4.3.5}
\]

When time is 0, the term \((1 - e^{-\frac{t}{\tau}})\) is \((1 - e^{0/\tau})\), or 0, and the response equals the level 1 response. When time is infinity, the term \((1 - e^{-\frac{t}{\tau}})\) is 1.0, and the response equals the level 2 response.

### 4.3.4 Second Order Equations

*Mathematics possesses not only truth, but supreme beauty – a beauty cold and austere, like that of sculpture, without appeal to any part of our weaker nature, sublimely pure, and capable of a stern perfection such as only the greatest art can show.*

- Bertrand Russell

Some biologically-important differential equations contain second derivatives, $\frac{d^2 y}{dx^2}$, and are called second order differential equations. A simple example of these is:

\[
x + \frac{1}{A \omega^2} \frac{d^2 x}{dt^2} = 0 \tag{4.3.6}
\]

The second derivative of a sine or cosine also contains a sine or cosine term:

\[
x = A \sin \omega t \tag{4.3.7a}
\]
\[
\frac{dx}{dt} = A \omega \cos \omega t \quad \text{(4.3.7b)}
\]

\[
\frac{d^2x}{dt^2} = -A \omega^2 \sin \omega t \quad \text{(4.3.7c)}
\]

So, \( x = A \sin \omega t \) can be seen to satisfy equation 4.3.6. Biological responses displaying this type of behavior are periodic in nature.

Figure 4.3.2. Sine and cosine waves.

4.3.5 Periodicity

We often forget that the burden of the Biological Engineer is to have some level of understanding of all of biology, just as we must be familiar with general engineering approaches to problem-solving.  

-Raj Tonnash
Some biological responses are periodic, varying in a predictable manner between two limits. The firing of certain nerve cells and the levels of circulating hormones are two examples of periodic responses. Another example is predator-prey population dynamics (see also Section 6.20.3).

Oscillatory behavior (Figure 4.3.2) is described mathematically by combinations of sines and cosines:

\[
\text{response} = (\text{magnitude})(\sin \frac{2\pi t}{T}) \tag{4.3.8a}
\]

\[
\text{response} = (\text{magnitude})(\cos \frac{2\pi t}{T}) \tag{4.3.8b}
\]

Figure 4.3.3. Fundamental and harmonics.
where $t$ is time, and $T$ is the period of the response. The difference between the response in equation (4.3.8a) and that in equation (4.3.8b) is that a sine

Figure 4.3.4. A complex waveform expressed as the sum of sines and cosines. At the top is a ramp function expressed as the sum (shown as a dotted line) of the fundamental, second, and third harmonics (shown individually as solid lines). At the bottom is the same function composed of ten harmonics. The more harmonics are used, the closer will be the representation to the actual waveform.
wave varies from zero to a positive magnitude to zero to a negative magnitude, and returns to zero. The cosine wave appears to have the same shape except that it begins and ends at the positive magnitude rather than zero. Inverting the period $T$ gives the frequency of the signal. The frequency of a simple sine or cosine wave is called the fundamental frequency. Frequencies related to the fundamental frequency by integer multiples are called harmonic frequencies (Figure 4.3.3). If the system is nonlinear, as many biological systems are, they may generate subharmonics.

More complex periodic responses require the sums or differences of sines and cosines with different periods. Fourier series is a means to express any periodic response in terms of the sum of sine and cosine waves with fundamental and harmonic frequencies (Figure 4.3.4).

### 4.3.6 Nonlinear and Nonconstant Equations

*Round numbers are always false.* - Samuel Johnson

The previous first and second order differential equations were extremely simple. They had derivatives that were raised only to the first power, they had constant coefficients, and they had only one independent variable, time. Not all equations used to describe physical or biological phenomena are so simple.

Many biological happenings are nonlinear. They may oscillate, but not with any set frequency. They may form exponential-like responses, but cannot be characterized by one time constant. Input-output relationships may not follow idealized forms. In these cases, the biological engineer must either resort to nonlinear equations or to numerical solutions to describe these phenomena.

A case in point is adaptive control systems, to be described in the next Section. Most control systems are based upon first and second order differential equations, but adaptive systems can change their responses to satisfy special requirements. Biological systems are particularly adept at this: they can often change the type or magnitude of response when simple predetermined responses are no longer adequate. Systems of this sort can be described by differential equations the coefficients of which are themselves dependent on magnitudes of the variables. The general approach is this:

1. determine by experiment the expected form of the response
2. choose an equation that adequately matches the essence of the response
3. incorporate the equation in a model that can be used to predict future responses.

See Section 1.4.
4.3.7 Integration

Nothing someone had ever measured was now or ever could be the same as before.  

-Daniel Kehlmann

An integral is an infinite sum of infinitesimal elements. Integration is the process of determining the value of an integral.  

Integrals are often used to determine areas under curves when the curve is determined by a known mathematical function. One example of where this is important is in the determination of the work required to move a fluid (air in the lungs or blood in the vasculature or cytoplasm in the cell).  

The volume flow rate \( \frac{dV}{dt} \) is often known over time. The pressure in the system is many times given as the pressure developed in a compliance element, \( C \), across a resistance element, \( R \), and across an inertance, \( I \) (see Section 2.1):

\[
p = \frac{dV}{dt} R + \frac{1}{C} \int \frac{dV}{dt} dt + I \frac{d^2V}{dt^2} \tag{4.3.9}
\]

The rate of work, \( \frac{dW}{dt} \), is \( p \frac{dV}{dt} \), so the amount of work is:

\[
W = \int \frac{dW}{dt} dt = \int p \frac{dV}{dt} dt = \int \left[ R \left( \frac{dV}{dt} \right)^2 + \frac{1}{C} \frac{dV}{dt} \int \frac{dV}{dt} dt + I \frac{dV}{dt} \frac{d^2V}{dt^2} \right] dt
\]

(4.3.10)

where the symbol \( \int \) denotes integration.

Example 4.3.1 Human Population of the World

The world population in 1980 was about 4.432 billion and growing at a rate of 1.7% (Alocilja, 2002). Modeling world population growth as an exponential process,

\[ N(t) = (4.432 \times 10^9) e^{0.017t} \]

where \( N(t) = \) world population at any time (t, years after 1980). Calculating the world population in 1986, gives:

\[ N(1986) = (4.432 \times 10^9) e^{0.017 \cdot 6} = 4.908 \text{ billion people.} \]
A census of people gives the world population in 1986 to be 4.9 billion people. Thus, the exponential model of world population appears to be a good fit. Exponential growth is also seen to be valid for microbes, plants, birds, insects, cancer cells, or any biological entity not limited by environmental resources.

Example 4.3.2 Classroom Ventilation

Studies of indoor air quality necessarily include ventilation of occupied spaces. Inadequate ventilation is often the cause of sick building syndrome, wherein occupants complain of various ailments associated with their presence inside the building.

Ventilation efficiency can be characterized by measured levels of metabolically-generated CO₂ or by the concentration decay of a passive tracer gas. Sulfur hexafluoride (SF₆) is often used for this purpose, and it is introduced into the space under consideration. The faster the decline of SF₆, the greater the ventilation of the space.

Bartlett et al (2004) have modeled CO₂ in naturally-ventilated classrooms occupied by children. The only air exchange present in naturally-ventilated classrooms is provided by infiltration and exfiltration, mostly through open doors and windows. Bartlett et al (2004) gave the basic mass balance equation for CO₂ as:

\[ a(t)[c(t) - c_{out}] + b(n(t)) = \frac{dc(t)}{dt} \]

where
- \( c(t) \) is the classroom concentration of CO₂
- \( c_{out} \) is the concentration of CO₂ in the atmosphere surrounding the classroom
- \( a(t) \) is the air exchange rate
- \( b \) is the CO₂ generation rate, dependent upon \( n(t) \), the number of people in the room.

All parameters with “(t)” appended can vary with time.

Considering this first-order differential equation in light of the mass balance (see Section 2.2), we see that the terms in the above equation correspond to:

\( (\text{rate of CO}_2 \text{ in} - \text{rate of CO}_2 \text{ out}) + \text{rate of CO}_2 \text{ generated} = \text{rate of CO}_2 \text{ stored} \)

And, by introducing the term:
\[ \Theta(t) = (c(t) - c_{\text{out}}) / c_{\text{out}} \]

they transformed the above mass-balance equation into:

\[ \frac{d\Theta}{dt} + \alpha(t)\Theta = \beta(t) \]

which is formed from dimensionless terms and has the advantage that the variable \( \Theta \) does not carry a specific set of units.

Note that this last equation would be of the standard form of a first-order differential equation, with an exponential function solution, except for the time dependencies of the terms \( \alpha(t) \) and \( \beta(t) \). These terms make this a nonlinear differential equation that most likely must be solved numerically.

**Equation 4.3.3 Respiratory Work Rate**

The work of breathing represents an energy drain on the exercising human or animal. About 8-10% of the body’s oxygen consumption is spent for respiration during heavy exercise (Johnson, 1991). Many adjustments are made in the respiratory system to make respiration more efficient. Among them are the breathing airflow waveshape, which changes from sinusoidal at rest to trapezoidal during exercise.

The work of breathing can be calculated with a suitable model of the respiratory system. Resistance, compliance, and inertance are all present. Thus, the pressure that must be developed by the respiratory muscles is (Johnson, 1993):

\[ p = R\dot{V} + V/C + I\ddot{V} \]

where  
\( p = \) pressure  
\( R = \) resistance  
\( C = \) compliance  
\( I = \) inertance  
\( V = \) lung volume  
\( \dot{V} = \) volume flow rate  
\( \ddot{V} = \) volume acceleration

Assuming a sinusoidal inhalation flow waveshape,

\[ \dot{V} = \dot{V}_{\text{max}} \sin \frac{\pi t}{T} \quad 0 \leq t \leq T \]
where  
\( t = \text{time} \)  
\( T = \text{time for inhalation to occur} \)  
\( V_{\text{max}} = \text{peak flow rate} \)

Lung volume can be obtained by integrating flow rate over time:

\[
V = \int_{0}^{T} \dot{V} \, dt = \int_{0}^{T} \dot{V}_{\text{max}} \sin \frac{\pi t}{T} \, dt = V_{0} + \frac{\dot{V}_{\text{max}} T}{\pi} \left(1 - \cos \frac{\pi t}{T}\right)
\]

where \( V_{0} = \text{initial lung volume} \).

Volume acceleration can be determined by differentiating flow rate:

\[
\ddot{V} = \frac{d}{dt} \left( \dot{V}_{\text{max}} \sin \frac{\pi t}{T} \right) = \frac{\dot{V}_{\text{max}} \pi}{T} \cos \frac{\pi t}{T}
\]

Work can be found by integrating the product of pressure and flow rate over time. Average work rate is the total work divided by inhalation time:

\[
\dot{W} = \frac{1}{T} \int_{0}^{T} p \ddot{V} \, dt = \frac{1}{T} \int_{0}^{T} \left(R \ddot{V} + \frac{\dot{V}}{C} + I \dot{V}\right) \dot{V} \, dt
\]

\[
= \frac{1}{T} \int_{0}^{T} \left(R \ddot{V}^2 + \frac{\dot{V} \ddot{V}}{C} + I \dot{V} \ddot{V}\right) \, dt
\]

\[
= \frac{1}{T} \left\{ R \int_{0}^{T} \dot{V}_{\text{max}}^2 \sin^2 \frac{\pi t}{T} \, dt + \frac{1}{C} \int_{0}^{T} \left(\dot{V}_{\text{max}} \sin \frac{\pi t}{T}\right) \left(V_{0} + \frac{\dot{V}_{\text{max}} T}{\pi} \right) \left(1 - \cos \frac{\pi t}{T}\right) \, dt + I \int_{0}^{T} \left(\dot{V}_{\text{max}} \cos \frac{\pi t}{T}\right) \left(\dot{V}_{\text{max}} \sin \frac{\pi t}{T}\right) \, dt\right\}
\]

\[
= \frac{R \dot{V}_{\text{max}}^2}{2} + \frac{2 \dot{V}_{\text{max}}^2 T}{\pi^2 C} + \frac{2 \dot{V}_{\text{max}} V_{0}}{\pi C}
\]

Work rates for other waveshapes can be determined in a similar manner. The sinusoidal waveshape has been found to be 6% more costly than a trapezoidal waveshape during light exercise and 9% more costly during heavy exercise (Johnson, 1993).
Applications and Predictions

1. The amount of drug remaining in a biological system will follow a decreasing exponential relationship.
2. The rate of growth of an individual can be described as an increasing exponential curve followed by a decreasing exponential curve.
3. Temperature receptors will respond exponentially to a sudden change in temperature.
4. Movement of the eyes will be oscillatory.
5. Most biological responses will be exponential, not oscillatory.
6. Sleeping patterns will be periodic.
7. Disease patterns are often periodic, based upon a periodic environmental fluctuation.
8. Harmonic analysis can be used for diagnosis of heart problems.
9. Membrane activities can be modeled with exponential equations.
4.4 Control Systems

Because of the great complexity in a biological system, extended chains of mathematical reasoning are less relevant in biology than in physics and engineering. . .

-E. Körner and G. Matsumoto

Biological stability (called homeostasis) is achieved through active control. Whether we consider the intracellular production of enzymes or the maintenance of whole-body posture, it is important for a living system to be able to sense the level of the controlled variable and then respond in a manner to correct discrepancies between desired and actual levels.

All control systems require:
- sensors
- actuators
- controller
- means to communicate among these elements.

4.4.1 Sensors

When we have no control over our sense organs, we have no control over the world. We become a slave to it. At the beck and call of the world, we let our energies run dry, dissipating all vitality from our personality. What remains is but a carcass of the physical body: a mere biological unit moving about, with its physiological activities intact but with no personality to assert, plan, or achieve. If a society is made up of such exhausted and empty human beings, no scientist can help improve it, no politician can save it, no economist can develop it.

-Swami Chinmayananda

Any control system that purports to respond to environmental stimuli must sense those stimuli. It may seem overly simple to realize that an environmental attribute does not really exist for that biological unit if the attribute cannot be sensed. Thus, humans cannot see ultraviolet radiation the way honey bees can, they cannot hear high frequency sound the way bats can, they cannot sense magnetic fields the way migrating birds can, and they are not aware of electrostatic fields as sharks are. To humans, lack of sensation indicates no information available, and we are not even aware of our unawareness. Signals outside our range of sensation do not exist for us.

Sensors take several forms (see Sections 6.19 and 6.20). Sensors, called receptors in biology, are transducers that change one type of signal into another more easily manipulated or communicated. Animal nervous systems are specialized for communication, so many animal receptors are those that convert chemical, mechanical, or electromagnetic radiation into a form that can be transmitted by neurons.
Often the most critical element in control systems designed to be used by humans is at the sensor level. Sensors must be reproducible and stable over very long periods of time, and these two criteria are difficult to meet. Indeed, one of the biggest impediments of controlled drug delivery systems, as for insulin in diabetics, has been the inability to produce a reliable and reproducible glucose sensor. Biological receptors are often nonlinear; their output signals are not often linear representations of their environmental input signals. In many of the feedback control systems to be subsequently described, this nonlinearity is not a shortcoming. Rather, it is hardly noticed.

Biological receptors are also often sensitive to the rate of change of the stimulus as well as to the stimulus level. Receptor outputs, therefore, often have two thoroughly mixed components representing the level of the stimulus and the rate of change of that level. As we will see, rate of change information can convey an advantage to the biological system.

How Receptors Work

There are many kinds of receptors in the body, and several different kinds of mechanisms that transform the adequate stimuli into a series of action potentials. For those receptors that receive energy in some form (for example, light, mechanical energy, or heat), the energy can be used directly to change the resting potential of the cell. It would seem likely that the energy would somehow upset the gel structure of the cell (Pollack, 2001) by agitation of the straight protein strands that form its foundation. Structured water between the strands would be disrupted and allow Na\(^+\) ions to flood into the cell, at least locally. This will change the cell resting potential (this change is called the *generator potential*), moving the cell potential closer to the threshold potential of the neuron, and making it easier to fire an action potential. In the case of vision sensors, the addition of light actually decreases the frequency of action potentials (Zimmerman, 1995).

Some chemicals can pass into a receptor cell and either bind to specific receptors inside the cell or modify the formation of mRNA from DNA. However, most chemical receptors use a complex cascade known as a second messenger system. Most signal molecules are water soluble and too large to pass freely through the plasma membrane of the cell. In the surface of the membrane are located fixed receptor proteins, and these are shaped to be able to accommodate smaller molecules, called ligands or agonists, like a peg in a hole of complementary shape (see figure). Depending on the shape of the receptor protein, the receptor
response may be more or less specific to a type of chemical. When the mating of receptor protein and ligand takes place, there is a conformal change in the protein.

Inside the membrane is one of a number of purine (a double-ringed organic structure) molecules based upon guanine (the same nucleotide base that helps form DNA and RNA). These are called G-proteins.

G-proteins can form complexes with phosphate compounds similar to ATP. The two complexes of importance here are GDP (two phosphates) and GTP (three phosphates). When GDP is bound to the G-protein, it is inactive; nothing happens; this is the resting state. When GTP is bound to the G-protein, then a series of actions follows.

Second messenger receptors. When the agonist (A) binds to the receptor site (R), the receptor loses its affinity for GDP and instead binds GTP. When the GTP binds to the G protein (α, β, γ), the G protein subsequently dissociates and fixes to the target enzyme (E). The target enzyme is either activated or inhibited. Hydrolysis of GTP to GDP causes the Gα-GDP complex to lose its affinity for the target enzyme and it returns to the receptor site. The agonist leaves, returning the receptor to its resting state (Sleight and Lieberman, 1995).
Surfaces of cell membranes abound with biomolecules called ligands that function normally as outer membrane receptors (Figure 4.4.1) or other structural features (Kim et al, 2004). Ligands act as links between the cell and external entities such as viruses, cells, and other bioactive molecules. These ligands are chemically tethered to surfaces by long (approximately 100 Å) organic linking molecules. Ligands are important for pathogens to be recognized by the cell, and ligands are important for the cell to be recognized by pathogens. It is this property that can be useful in the design of biosensors. Capture events can be detected almost immediately by intrinsic fluorescence of microbes, toxins, or DNA. There is no waiting for microbial growth necessary for culture plate counts, or for enzyme reactions used in ELISA (see Section 3.5), or for replication of DNA used in PCR (see Section 5.3.4). As long as the fluorescence can be detected, the result is real-time pathogen or contaminant detection.
Figure 4.4.1. The surface of a cell is rough, and includes many molecules that act as receptors for other substances in the surrounding fluid.

Example 4.4.1. Making Bitter Food Taste Better

The tongue is home to more than 10,000 taste buds. Each of these detects specific chemicals in food and drink. There are taste receptors for sweet, salty, bitter, savory, and sour-tasting chemicals.

Not all people taste food in the same way. For some people, very sensitive bitter receptors make certain foods unpleasant to taste.

Aside from avoiding bitter foods, the only thing that has been done to make these foods more palatable is to add salt and sugar to mask the bitter flavors. Now, new chemicals can be added to the food to block bitter taste receptors. With knowledge about how these receptors work, specific biochemicals with correct configurations can keep bitter flavors from triggering responses from their receptors (Lashinsky, 2007). Blocking receptors from responding to stimuli is one way to manage unpleasant or painful perceptions.

4.4.2 Actuators

*Information is physical.* - Rolf Landauer

Actuators are the means for a control system to have an effect. Actuators may be local and mechanical, as are muscles, or they may be diffused and chemical, as are glucose respiration processes in cells.

Many of the more visible control actions are concerned with movement. Many microbes and most animals are capable of locomotion (see Sections 2.9, 6.9, and 7.3.9). The actuators in these cases are the two protein fibrils actin and myosin that slide past each other and shorten to produce movement. On a much longer time and space scale, plant populations move in response to climate changes, and, in this case, the actuators are the seeds that disperse in various ways.
Muscle Types

The contractions of all muscle cells depend on two proteins that appeared in very early life: actin and myosin. Actin and myosin filaments, in units called sarcomeres, slide past each other when a propagating phase change occurs.

Actin can undergo a phase change from an extended state to a folded state. The surrounding water forms a gel structure in the extended state. It is thought that this or a similar phase change moves the actin filaments along the sarcomere. When actin moves, it displaces in multiples of the 2.7 nm cross-link spacing. Three different types of muscle filaments are capable of shortening (Pollack, 2001).
Muscle Types cont.

Sarcomeres are the basic muscle units.
**Muscle Types** cont.

*Skeletal muscles* contract when depolarized by neural signals transmitted at the neuromuscular junction. Each muscle fiber has one synaptic connection, but each motor neuron typically branches and controls several muscle fibers. The larger and more powerful muscles are controlled by neurons branched many times. The contractile apparatus composed of motor neuron and all the fibers it controls is called a *motor unit*.

Organization of muscular motor units (Campbell et al, 1999).
Muscle Types cont.

There are fast twitch and slow twitch skeletal muscle fibers specialized for different uses. Fast fibers are used for rapid, powerful contractions. They do not fatigue quickly in the absence of oxygen; they utilize anaerobic metabolism. Slow fibers are meant to sustain long periods of repeated contractions without fatiguing; however, they do require a steady supply of oxygen in order to maintain this action. They have many mitochondria, a rich blood supply, and an oxygen-storing protein called myoglobin. Fast twitch fibers are used for strong voluntary muscular movements such as running or flying; slow twitch fibers are used to maintain posture or to carry heavy weights.

Cardiac muscle (myocardium) is only found in the heart. It has the ability to generate action potentials and to pass depolarization (and subsequent contraction) from cell to cell without neural intervention. Cardiac muscle otherwise has some of the same internal cellular organization as skeletal muscle.

The actin and myosin of smooth muscle is organized differently from the striated (containing sarcomeres) skeletal and cardiac muscle cells. Smooth muscle cannot generate as much tension as striated muscle, but it can contract over a wider range of lengths. Smooth muscle is mostly found in the walls of hollow vessels of the digestive system, respiratory system, and blood vessels. There it can control flow in and through the vessels as it contracts. Smooth muscle in the urinary bladder can produce a nearly constant pressure on the fluid for a wide range of bladder volumes.

Some invertebrates contain muscles similar in many ways, but different in others. The flight muscles of insects, for instance, can spontaneously depolarize and contract very rapidly (Campbell et al, 1999). Clam shell muscles can clamp the shells closed with a low energy consumption for as long as a month (Campbell et al, 1999).

4.4.3 Communications

When asked what distinguishes a functioning animal from an inviable heap of cells, most biologists would draw attention to the hormonal and nervous systems.

- Jonathan Cooke

Without some means to transfer information between control system elements, there can be no precise control. Intact eyes without an optic nerve to convey visual information to the brain are no better than no eyes at all. Cellular insulin receptors without access to blood flow cannot function as intended. Likewise, communication among individuals in an ecological
community is just as necessary as scouting reports to an army; together, the group can respond to the information obtained from the outside.

There are two broad classes of communication in a complex animal organism:

- neural
- humoral (chemical).

Distinguishing between these in the traditional ways, humoral communication generally uses circulating fluids as the medium of intercourse. The circulating blood in animals, cytoplasmic streaming in single cells, or phloem fluids in plants are examples of these. Neural communication uses the nervous system, and can deliver messages to specific locations. The nervous system has been thought to deliver its messages by membrane depolarization of specific target cells. Actually, the nervous system is a chemical delivery system that is very specifically targeted. Only in nerves that end on muscle cells is membrane depolarization of most importance. For other cells, it is the neurotransmitters that are released at the ends of the nerves that initiate action. Thus, both neural and humoral communications mechanisms turn out to be chemical. Humoral communication differs from neural communication in that it is more general throughout an organism, and it is often slower to respond.

Plants and single-celled organisms, of course, do not have nervous systems, so they rely on circulating chemicals, often hormones, to convey messages. This limits the speed at which they can respond to stimuli.

Ecological systems, also, do not communicate by nervous systems. However, there is an analogous system of individual-to-individual communication that enables targeted information transfer. A system of this kind can be as simple as a microbe on the skin of a human or as complex as a nation.

Communications mechanisms must be fast enough to enable the control system to respond in a timely manner. Humoral communications cannot be used for fast responses, and many unmyelinated nerve fibers are also too slow for many purposes. The organism that can transmit information and process it the fastest often has a survival and reproductive advantage, so some specialized communication means (such as nerve fibers in myelin sheathes) have developed.

Although they are often diagrammed as straight lines or wires, communications systems are much more complex, and often display characteristics of a transmission line (Figure 4.4.2). Transmission lines have resistance, capacity, and inertia elements that slow and degrade a signal. This is true for both neural and humoral signals.
Figure 4.4.2. Representation of a transmission line with resistance, capacity, and inertia elements. If a sudden change is made on the left hand side of the transmission line, the resistances degrade the signal and slow the flow variable, the capacity elements must be filled, and the inertia elements slow the rate of change. Thus, the transmission line delays the output from appearing, and degrades its sharpness.

Communications delays can sometimes cause instability in a control system. Cheyne-Stokes breathing, characterized by periods of apnea interspersed with maximal respiration (Figure 4.4.3), has been found to be attributed to an abnormal delay in the respiratory controller (Hornbein, 1981). Delays, for whatever reason, cause instability because, by the time corrective action can be taken, the stimulus has already changed significantly.

Figure 4.4.3. Cheyne-Stokes breathing, shown here as respiratory flow rate with time, is characterized by periods of intense breathing followed by no breathing at all. This condition is caused by a delay in the control system.

**Autonomic Nervous System**

The peripheral nervous system (those nerves outside the central nervous system) of vertebrates is made up of both sensory (afferent) and motor (efferent) nerves (Figure). Each of these includes a bundle of nerve cells that connect to different sensors and actuators in the body. Motor nerves can be either somatic (mostly voluntary) or autonomic (mostly involuntary).
The autonomic nervous system is composed of two groups of nerves that have mostly (although not always) antagonistic effects. The sympathetic system generally produces reactions to respond to emergencies or exigencies, the so-called “fight or flight reactions”. The parasympathetic system promotes homeostasis, calm, and routine activities. Parasympathetic nerves originate in the brain or the sacral region of the spinal cord. They supply visceral structures in the face and head through the oculomotor, facial, and glossopharyngeal nerves. The thorax and upper abdomen is supplied by the vagus nerve. The pelvic viscera are supplied by the pelvic branches of the second to fourth sacral spinal nerves. Parasympathetic neurotransmission uses acetylcholine, and are thus called cholinergic nerves.

Sympathetic nerves form ganglia (structures containing neuronal cell bodies) outside the spinal cord. Sympathetic neurotransmission uses norepinephrine, and are called adrenergic.
Autonomic Nervous System cont.

The autonomic nervous system (Campbell et al, 1999).
Action Potentials

Unlike epithelial and other types of cells that maintain a steady cell resting potential (see box, Section 3.8), the internal voltages of neurons and muscle cells can be altered by external events. In so doing, information can be transmitted along the cell. Neurons are particularly suited for this type of information transmission. They have relatively large cell bodies with branching, fiber-like extensions called dendrites to receive signals, and a long extension, called an axon, to transmit signals (Figure). The dendrites are located close to the axon terminals of one or oftentimes many other neurons; they function by receiving signals from the other neurons and integrating them in summation fashion. Some of these inputs may be excitatory, and some inhibitory; the dendrites receive these signals and their cellular potentials vary depending upon how many of each they receive within a short time.

The axon functions as a transmission line, bypassing the slow or unresponsive tissue between the cell body and the terminus of the axon. It is along the axon that the signal is transmitted. Some axons are short (in the central nervous system, or CNS), while some are very long. Some axons of the sciatic nerve extend a meter or more from the lower part of the spinal cord to the lower leg and foot. Here, the word “nerve” refers to a bundle of neurons, or nerve fibers, in a common sheath of connective tissue.
The signal itself is a cell potential depolarization that travels in wavelike fashion along the axon. The cell resting potential of about $-70$ mv actually reverses polarity to become $+30$ to $+35$ mv for a very short time. This is called the action potential, and it moves from one end of the axon to the other (Figure). The action potential is an all-or-nothing event, similar to flushing a toilet; once a threshold voltage is reached, the action potential starts and can’t be stopped until completion. Movement of the action potential is like the crowd doing the wave in a stadium, or like the peristaltic movement along the bowel.

Behind the depolarization is a repolarization that actually hyperpolarizes the cell potential for a while. The action potential normally moves from cell body toward the axon terminus, but it could be propagated in the other direction. The hyperpolarization following the action potential, however, eliminates the possibility of back-propagation because it moves the cell potential farther from the threshold voltage. This results in a latent (refractory) period during which the cell cannot easily form an action potential. The strength-duration curve shows the transmembrane current required to initiate an action potential following a previous action.
Action Potentials cont.

potential (Figure). Rheobase is the minimum cell current that would initiate the action potential. Chronaxie is the time following an action potential for the current required to initiate the next action potential to be twice the rheobase value (Cuervo, 1976).

![Graph showing the relationship between current strength and time with Rheobase and Chronaxie marked.]

The current strength required to initiate an action potential decreases with time following the previous action potential (Cuervo, 1976).

So how does the cell potential reach the threshold voltage? The junction between two neurons is called a synapse, and this junction is usually located at the terminus of a presynaptic neuron and the axon hillock of the postsynaptic neuron. The axon hillock is the location where the axon joins the neuronal cell body, and it is here that the action potential begins. A signal coming from the presynaptic neuron can trigger an action potential in the postsynaptic neuron by causing the cell potential of the postsynaptic neuron to reach threshold voltage.

There may be connections with several presynaptic neurons, and not all of these lead to action potential generation in the postsynaptic neuron (Figure). Those connections that favor depolarization (moving toward threshold voltage) are called excitatory postsynaptic potentials (EPSP); those connections that tend to hyperpolarize (move the cell potential farther from the threshold voltage) are called inhibitory postsynaptic potentials (IPSP). There may be both temporal summation
and spatial summation of these postsynaptic potentials (Figure). If incoming signals occur from one presynaptic neuron fast enough, the signals can sum to have a larger effect than just one signal would have. If in this instance we consider just EPSPs, then this example of temporal summation will make easier the formation of an action potential in the postsynaptic neuron. Similarly, if signals from two or more presynaptic neurons arrive nearly simultaneously, they can sum to produce a larger effect than if they arrived separately.

Although they cannot form an action potential, the dendrites can influence the cell potential at the axon hillock, and thus raise or lower the neuronal sensitivity to action potential formation. They do this by summing other inputs from other neurons that, again, may be either excitatory or inhibitory.

Thus, the action potential is formed from spatial and temporal summation of EPSP and IPSP signals with graded input from the
Action Potentials cont.

Both temporal and spatial summation of subthreshold potentials are possible (Campbell et al, 1999).

dendrites. This makes it possible that neurons can form complex responses from relatively simple input signals.

Axons exhibit all the electrical characteristics of a transmission line. In particular, they offer resistance to current flow during the action potential, and, because there is a separation between unlike charges inside the cell and outside the cell, they exhibit capacitance as well. You will recall that in the resting state there are more negative charges inside the cell than outside, and during the action potential there are more positive charges inside the cell than outside. Supplying the current through a resistor to reverse the charge on a capacitor takes some time, and follows an exponential time course (see Section 4.3). The time constant of the exponential curve is given by the product of resistance and capacitance. The rate at which action potentials can be propagated along an axon is limited by the time constant of the axon.

For very long axons, especially those located outside the CNS, this time is too slow for adequate information transfer, so reducing the time constant would be of great survival benefit. Resistance can be reduced by increasing the size of the axon, because resistance is inversely related to the cross-sectional area through which current flows. Transmission speed can be increased from several centimeters per second in very thin axons to about 100 m/sec in the giant axons of invertebrates such as the squid and lobster (Campbell et al, 1999).

There are, also, means to reduce capacitance of the axon, and the
Action Potentials cont.

means used in many vertebrate axons is to increase the distance between charges. Peripheral axons are enclosed in myelin sheaths formed by supporting cells called Schwann cells; some larger CNS axons are enclosed in myelin sheaths formed by oligodendrocyte cells. Myelin is a fatty substance that acts as an electrical insulator preventing current flow.

There are breaks in the sheath, between supporting cells, called Nodes of Ranvier. These allow for currents to flow to support propagation of the action potential. The action potential, and associated currents, proceed by jumping from one node to the next (called saltatory conduction) at a rate of up to 150 m/sec.

Until this point, we have avoided details about the nature of the current flow and mechanisms causing the action potential. This is because certain mechanisms are generally accepted as explanations for the action potential, but, related to establishment of the cell resting potential (see box, Section 3.8), there are alternative theories.

Sodium, potassium, chloride, and calcium ions carry the currents associated with the action potential. The two that have seemed to be most important are Na\(^+\) and K\(^-\). According to the traditional explanation, the cell resting potential was maintained by a Na\(^+\)-K\(^-\) ion pump (Figure).

![Diagram of sodium and potassium movement across a cell membrane. Passive diffusion occurs because of concentration gradients across the membrane. A sodium-potassium pump, powered by ATP, moves sodium and potassium against their concentration gradients (Campbell et al, 1999).](image-url)
Action Potentials cont.

When the threshold voltage was reached, or the membrane was disturbed in some way, Na⁺ conductance of the membrane increased, which allowed Na⁺ to flood into the cell. This reversed the cell potential, and formed the positive-going portion of the action potential (Campbell et al, 1999). A slower increase in K⁺ conductance allowed K⁺ to flow out of the membrane. This, and the flow of Cl⁻ into the cell, brought the cell potential back down, forming the trailing negative-going part of the action potential. The refractory period following the action potential was caused when the ion pump reestablished inside and outside Na⁺ and K⁺ concentrations.

Pollack (2001) has a different explanation, one that is based more on the physicochemical basis of the cell, and one that explains newer research observations, but one that does not have an explanation for as many details as the traditional hypothesis. You may recall that the cell membrane is very porous, and its integrity is not essential to the presence of the cell resting potential (see box, Section 3.8). The resting potential comes about because of the net negative charge on immobile proteins inside the cell, and the structured water (gel) that surrounds these proteins. There is hardly any room for included ions, especially large hydrated ions such as Na⁺. Pollack (2001) presents evidence that the action potential requires an intact cellular cytoskeleton. The cytoskeleton is a dense polymer-gel matrix composed mainly of cross-linked actin (a contractile protein present in muscles of all animals from protozoa to vertebrates, and in the microfilaments of all cells) and microtubules (hollow filaments 20-25 nm in diameter found in eukaryotic cells and composed of an actin-like protein called tubulin). The cytoskeleton in nerve cells is present just below the membrane and carries a high negative surface charge. According to Pollack, a phase change in the cytoskeleton during the action potential, accompanied by water absorption and heat liberation, disrupts the structured layers of water in the gel. The liberation of heat would be expected if a highly ordered structure, which takes energy to form, were to become less ordered.

Na⁺ would be expected to rush into the unstructured water because of a concentration difference between Na⁺ outside the cell and Na⁺ inside the cell. However, it has been observed that action potentials can be generated even when Na⁺ and K⁺ are absent (Pollack, 2001). Ca²⁺, however, must be present.

Ca²⁺ is a small divalent ion that can cross-link with different protein strands and hold them together. If Ca²⁺ is displaced by Na⁺, a monovalent ion that cannot cross-link protein strands, then the protein structure that holds water in place cannot be maintained. Thus, as Na⁺ rushes into the cell, it may displace Ca²⁺, and add to the tendency of the gel to become less ordered. Water also rushes in, and the system of protein strands expands. Only when the strands are pushed apart far enough will there be enough...
recoil to reverse the process. Thus, the action potential is a physicochemical process involving the expansion and contraction of the cytoskeleton matrix.

What can initiate this action? Catecholamines are chemicals formed from a benzene ring, adjacent hydroxyl groups, and an amine group (tyrosine). Examples are epinephrine (adrenalin), dopamine, and norepinephrine. They have specific effects on the nervous system.

Catecholamines are necessary for the repetitive action potentials generated by heart pacemaker cells. Binding of catecholamines in the cell could be the mechanism by which the action potential is initiated.

Indeed, the transmission of a neural signal at the synapse is usually by chemical means (some direct electrical connections are present in the giant axons of crustaceans and fishes, but these are rare). At the end of the axon are numerous sacs called synaptic vesicles, each of which contains small amounts (but thousands of molecules) of neurotransmitter substances. When the depolarization of the action potential reaches the synapse, neurotransmitter is released into the small gap between the neurons.

### Neurotransmitter Substances (Campbell et al, 1999)

<table>
<thead>
<tr>
<th>NEUROTRANSMITTER</th>
<th>STRUCTURE</th>
<th>FUNCTIONAL CLASS</th>
<th>SECRETION SITES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetylcholine</td>
<td><img src="image" alt="Acetylcholine Structure" /></td>
<td>Excitatory to voluntary skeletal muscles, excitatory or inhibitory at other sites</td>
<td>CNS, PNS: vertebrate neuromuscular junction</td>
</tr>
<tr>
<td>Biogenic Amines</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norepinephrine</td>
<td><img src="image" alt="Norepinephrine Structure" /></td>
<td>Excitatory or inhibitory</td>
<td>CNS, PNS</td>
</tr>
<tr>
<td>Dopamine</td>
<td><img src="image" alt="Dopamine Structure" /></td>
<td>Generally excitatory; may be inhibitory at some sites</td>
<td>CNS, PNS</td>
</tr>
<tr>
<td>Serotonin</td>
<td><img src="image" alt="Serotonin Structure" /></td>
<td>Generally inhibitory</td>
<td>CNS</td>
</tr>
<tr>
<td>Amino Acids</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GABA (gamma-aminobutyric acid)</td>
<td><img src="image" alt="GABA Structure" /></td>
<td>Inhibitory</td>
<td>CNS, invertebrate neuromuscular junction</td>
</tr>
<tr>
<td>Glutamine</td>
<td><img src="image" alt="Glutamine Structure" /></td>
<td>Excitatory</td>
<td>CNS, invertebrate neuromuscular junction</td>
</tr>
<tr>
<td>Aspartate</td>
<td><img src="image" alt="Aspartate Structure" /></td>
<td>Excitatory</td>
<td>CNS</td>
</tr>
<tr>
<td>Neuropeptides</td>
<td>Substance P</td>
<td><img src="image" alt="Substance P Structure" /></td>
<td>Excitatory</td>
</tr>
<tr>
<td>Met-enkephalin (β-endorphin)</td>
<td><img src="image" alt="Met-enkephalin Structure" /></td>
<td>Generally inhibitory</td>
<td>CNS</td>
</tr>
</tbody>
</table>
Receptors in the postsynaptic neuron sense the presence of the neurotransmitter, and may then initiate its own action potential.

There are several known neurotransmitters (Table). The most common of these is acetylcholine, which, as with many other neurotransmitters, may be either excitatory or inhibitory on the postsynaptic cell, depending on the type of receptors present on different cells. Some of these substances may be released into the bloodstream as hormones by endocrine glands. When acting as hormones they have much more general effects on the body.

Psychoactive drugs, including LSD and mescaline, produce their hallucinatory effects by binding to serotonin and dopamine receptors in the brain. Endorphins function as natural analgesics and decrease the perception of pain as well as produce euphoria.

It is important that neurotransmitters are destroyed soon after release. If they weren’t, their effects would be felt too long after they were intended. There are enzymes present in the synapses to hydrolyze neurotransmitters soon after they are released by the presynaptic neuron. Cholinesterase is the enzyme that acts on acetylcholine. A lack of cholinesterase would cause permanent depolarization of the postsynaptic neuron, and this neuron would then become ineffective for transmitting signals. Certain phosphate insecticides (and similar chemical agents) work in this way.

Brain neurotransmission (passing a signal from one neuron to the next) occurs within 0.3-100 msec (synaptic time delay) and over a distance of 30-50 nm (synaptic cleft) (Anonymous, 2007). These small scales of time and distance pose challenges for those intending to measure neural activities. New analytical methods are constantly being sought for real-time, small-scale measurements in biological systems.

4.4.4 Closed-Loop Feedback Systems

Much of human behavior is hard-wired. But, unlike the heart, liver, or even our genes, the brain can respond in a dynamic way not only to internal physiological cues but also to unpredictable external ones, and it can embody that response in future behavior. —Judy Illes

We can conceptualize this control system as a loop (Figure 4.4.4). The input level is sensed, the control system acts on the input level, and this, in turn, changes the input level. This type of control system is called feedback.
control, because information about the output of the system is fed back to the input in order to maintain stability.

Feedback usually gives very good control. The input level is sensed and compared to a set-point level. The difference between these two is then used to form a correction at the output, and this changes the input level. This is called proportional control.

There are times when this feedback can lead to oscillations, especially if there are delays discussed earlier. Oscillations in feedback control systems represent loss of control because they are difficult to stop once started. To improve stability, we could add rate of change information somewhere in the control system. A system that responds not only to the input level but also to the rate of change of the input anticipates where the input level will be and corrects ahead of time. Perhaps that is why so many biological receptors are sensitive to the rate of change in addition to the actual level of input.

We have just described proportional (P) control and proportional plus derivative (PD) control. Integration can be added to a controller, which gives it reset action, but also can exacerbate instability. There are proportional plus integral (PI) and proportional-integral-derivative (PID) controllers. These classical types are used where the system dynamics (the Plant) are well defined.

It is very difficult to tell the differences in the actions of the different kinds of controllers as long as the feedback loop is unbroken. Differences are only apparent when the loop is opened for some reason.

The box labeled “Plant” in Figure 4.4.3 is the process to be controlled. An example of what we mean might be the insulin control system (see Section 6.20), where the input signal relates to the rising glucose level of the blood and the beta cells of the pancreas are the sensors. Controller action takes

![Figure 4.4.4. Generalized feedback control loop. A portion of the output signal is fed back to the controller input, where it is subtracted from the reference level input. The controller then acts upon this difference, feeding it to the process to be controlled, known as the plant. The feedback signal may be conditioned before being compared with the reference input.](image)
place within the cells, and insulin release is stimulated. The Plant in this system is represented by the glucose uptake in the body cells and liver due to increased circulating insulin.

4.4.5 Open Loop Systems

*Make an estimate before every calculation, try a simple physical argument before every derivation, guess the answer to every puzzle.*

- John Wheeler and Edwin Taylor

An alternative to feedback control is open-loop control, where there is no loop. System output is determined by dead reckoning and there is no means to correct the output should there be an error. Open-loop control is faster, and requires less effort, than feedback control. Open-loop control works well for repetitive situations where there is small likelihood that an error will occur. An example of this is the signals sent to the leg muscles for walking on flat terrain. Only when the ground becomes rough or sloped is feedback normally required.

---

**Artificial Neural Networks (ANN)**

A human brain continually receives input signals from many sources and processes them in parallel to create appropriate output responses. There are billions of neurons in the brain that interconnect in a myriad of ways to form elaborate neural networks. ANN are an attempt to process information efficiently and quickly using brain neural networks as a model. Like brain neural networks, ANN have many neuron-like nodes that interact with one another; like brain neural networks, ANN must undergo a learning process before they are ready to process information automatically; like brain neural networks, ANN take information from a number of primary inputs and form useful outputs.

ANN architecture consists of an input layer of nodes, an output layer, and, sometimes one or more hidden layers between the two (Figure). The input layer receives (analog, not digital) information from a set of primary transducers. For example, different temperature or pressure measurements, or size information, may represent the input data. That data is fed to the ANN input layer.

The output layer typically consists of one node. The data value at the output node is the parameter that the ANN is meant to estimate.
Artificial Neural Networks (ANN), cont.

There can be one or more hidden layers, and each hidden layer can be composed of many different nodes. The number of hidden layers and the number of nodes per layer is empirically determined based upon performance. There is no theoretical basis for choosing the number of layers or the number of nodes. Frequently, one hidden layer is all that is required. The number of nodes for the hidden layer is often about twice the number of input nodes.

Although the ANN is given conceptually in the Figure, the actual ANN is usually a computer program without a physical basis. There are many ANN programs available, and differences among them are related to sizes and computational procedures.

Connections among the nodes are shown in the Figure. Each connection is given a weight and a threshold (or bias) value. These are estimated at the beginning.

The ANN is calibrated through a learning procedure where known input values are applied at the input layer and the output value at the output layer is computed. The computed output is compared to the ideal known value, and the weights and biases for each connection are adjusted to reduce the error. There are various computer algorithms available to make these adjustments efficiently. After a process of iteration, with connection weights and biases undergoing successive
Artificial Neural Networks (ANN), cont.

adjustment, the ANN computed output is nearly the same as the ideal output. At this point, again determined empirically, the learning process is stopped in order to avoid over-learning, where the ANN estimates the output for test conditions almost perfectly, but errors for other conditions become large.

After the learning step, the ANN is used to estimate output values for actual input data. ANN weights and biases are fixed during this process, and the ANN acts as an open-loop feedforward estimator.

4.4.6 Closed Loop Feedforward Systems

There are no creeds in mathematics. -Peter Drucker

There are times when a living system can anticipate that a corrective change will be required. Such a case might arise for a person living in a hot climate. Sweating can begin before a rise in body temperature because the body has acclimatized to the heat, and the thermoregulatory system knows that sweating will be required to maintain correct body temperature. When the response begins before a change in input level can be sensed, we call this feedforward (Figure 4.4.5). Feedforward does not form the same kind of loop

![Diagram of feedforward control](image)

Figure 4.4.5. Diagram of feedforward control. The feedforward signal is predicted from the input signal without waiting for slow feedback.)
as feedback and does not necessarily result in a very stable system. Indeed, feedforward can cause responses to be too extreme at times, but it also can be used to avoid catastrophic loss of control under extreme environmental circumstances, and, because it is anticipatory, the responses are very fast.

4.4.7 Adaptive Control Systems

The methods of computational biology are now so advanced that it’s conceivable to make a computer atlas of the nervous system and map the expression of hundreds of genes. -Chris Doe

An adaptive controller continually and automatically readjusts itself for proper operation in the presence of changing system dynamics or noise characteristics. It combines a parameter estimator and a control scheme that changes the control algorithm as needed. A block diagram of an adaptive system appears in Figure 4.4.6. Adaptive control is based on a linear differential equation with nonconstant coefficients, and is often used in drug-delivery systems (Woodruff, 1995) or where patient-to-patient variation is particularly wide.

Figure 4.4.6. Diagram of an adaptive control system with feedback. The difference between this system and the one in Figure 4.4.3 is the presence of the modifier that changes the controller performance.
4.4.8 Fuzzy Control Systems

Whatever a man prays for, he prays for a miracle. Every prayer reduces itself to this: “Great God, grant that twice two be not four.”

-Ivan Turgeniev

Fuzzy control systems are ones in which experts’ decision making rules are used to produce a control output. The rules may be mathematically based or not, and the controller output is usually sufficiently correct to perform the intended function. However, control with a Fuzzy system is not usually very precise.

Example 4.4.1 The Potted Rose (Alocilja, 2002)

A healthy rose plant will bloom with profuse, beautiful flowers. The plant requires water, so must be watered. If water is added without regard to the condition of the plant, this is an open-loop system. If water is added only when the plant needs it, or if the amount varies to meet plant moisture requirements, then the system is closed-loop with feedback. If extra water is added because you will be unable to care for the plant for a few days, then this is a feedforward control system.

Applications and Predictions

1. Learning will require feedback control; habit will allow open-loop control.
2. Cell growth will be regulated with feedback from its neighbors.
3. Removal of metabolic carbon dioxide during exercise can be considered to be feedforward control.
4. Movement of the limbs and joints will cause a direct and immediate rise in heart rate. This is feedforward control.
5. Hormone levels will be regulated by feedback control.
6. Enzyme-substrate inhibition will act as a feedback loop.
7. Going to the store to buy milk before it is all used up will be feedforward.
8. The stomach will produce gastric juice when hungry because of feedforward action.
9. Different control principles can be used to design artificial intelligence systems.
10. Myelinated axons will propagate signals faster than nonmyelinated axons.
11. Growth receptors in cells provide feedback to halt unchecked growth.
12. Feedback will regulate social interactions.
4.5 Optimization

*True science thrives best in glass houses, where everyone can look in.*

- Max Perutz

Maintenance of life and its many activities is a costly process. Energy is required to move, to maintain health, to compete for food, to reproduce, to grow, and to produce the chemicals of life. During normal times of homeostasis, the energy requirements of these processes are not particularly burdensome, but when additional demands are placed on the system, the organism may find that it cannot readily supply energy for all demands. Thus, adjustments may be made to minimize energy expenditure or to maximize output.

It is not exactly clear how optimization processes in living systems come about. However, mathematics may be used to find the optimum of a process if the mathematical description of that process is known over the range of interest. An optimum point can usually be described as a maximum or the minimum. We find that the rate of change of the variable of interest is zero at a maximum or minimum point. If the mathematical description is in equation form, then all we need do is to determine the rate of change of the variable of interest and set the rate of change to zero. Solving for the value of the variable of interest defines either the maximum or minimum point.

There are other tests we may use to determine if this is a maximum or a minimum. One of these is to know ahead of time that the point must be a maximum or must be a minimum, and that the place where the rate of change is zero has to be one or has to be the other, but cannot be either one or the other.

In graphical form, the place where the rate of change is zero is the point where the slope of the curve is zero (Figure 4.5.1). From this we can see that an optimum point may be found graphically where the slope is zero, mathematically where the rate of change is zero (using differential calculus), or numerically, where the optimum is estimated from known data values.

There may be local maxima or minima and global maxima or minima. The difference is that a local maximum or minimum is only the largest or smallest value over a very limited range. The global maximum or minimum is the most extreme value over the whole range. Biological data is usually simple and has only one maximum or minimum over the range of interest.

The maximum or minimum may only exist at one of the extremes of the range of interest. Graphically, this indicates that the highest or lowest value is at one end or other of the graph. If this situation happens with data from biological systems then it indicates that the process is not optimized by the biological system. In order for a process to be optimized, it must be able to be controlled, both from above the optimum or from below, and that measures must be able to be taken to move the process toward the optimum. If the optimum is at an extreme value, it cannot be controlled from both sides.
Figure 4.5.1. The optimum point is usually obtained where some dependent variable is minimized for some value of the independent variable. For instance, the rate of respiratory work is minimum at a particular breathing rate. A broad minimum, as seen here, means very little penalty is paid for operating away from the minimum.

An optimum point may be either shallow or deep. With shallow optima, the process may operate without significant penalty at a point away from the exact optimum. Thus, the system may be found to deviate one way or the other from the optimum, and a relatively large variation in the process operation is possible. With a deep optimum, there is a large penalty paid for deviations from the optimum, and the amount of variability of response is small.

Applications and Predictions

1. Many exercise parameters, for instance the rate of cycling, will be optimized to reduce energy expenditure.
2. When nutritious food becomes scarce, digestion will be optimized to extract as much nutrition as possible to minimize waste.
3. Kidneys save water when water is limited; they will excrete a lot of water when water is plentiful.
4. Animals and plants rest at night and are active during daylight.
5. Some animals hibernate when it is too cold or food is not available.
6. Whales reduce heart rate and metabolism when underwater.
7. The distribution of branches and leaves in a tree will optimize the use of sunlight.
8. Human hemoglobin is optimized to deliver the maximum amount of oxygen to the working muscles.
9. A neural cell network in the central nervous system becomes optimized to perform specific tasks.
10. Students eat cheap food to maximize nutrient content for the least cost.
11. Sizes of animal parts will be optimized to perform needed functions without expending too much energy.
12. A plant in the window will bend toward the light to maximize light reception.
13. Bacterial enzymes will adjust to food substrates available.
4.6 Information

*If you understand everything, you must be misinformed.*
- *Japanese Proverb*

In order to be able to act on and react to environmental conditions, biological systems must be able to obtain information from the environment, pass information internally to integrative centers where it can be processed, and then send information to actuation structures. Biological systems have developed special structures to obtain information, called sensors, and have developed special mechanisms to transmit this information internally.

There are many modes of information acquisition in biological systems. Information can take the form of chemicals either sensed or produced. In this case, environmental chemicals are sensed, and they may trigger production of other chemicals, and these other chemicals may be circulated throughout the organism, and other cells may use the circulating chemicals to produce other chemicals to react to the presence of the original chemicals in the environment.

The environment we mention here does not need to be external to an organism. It just needs to be external to the sensor.

Information can take the form of *neural* action potentials that travel from one end of a neuron to another, and from there across a synapse by chemical means to another neuron. These action potentials can be triggered originally by chemicals or by energy sources such as heat, light, or radiation.

In all information systems there is the problem of recognition of meaningful signals. *Noise*, or meaningless information, is present at all levels, but is especially important where the strength of the meaningful signals is the smallest. Biological sensors must be able to separate the signals from the noise. If not, then the system would expend much wasted energy reacting to meaningless inputs.

By the time a meaningful signal is recognized in some way, it is usually amplified by the system so that the noise becomes relatively smaller. At this point, any additional noise in the system may actually help the organism to maintain its awareness of the environment.

There is a thermodynamic limit to the smallest signal that can be sensed and distinguished from background noise. In general, the more sensitive the sensor, the more energy it takes to maintain that sensitivity. Thus, the most sensitive sensors are those related to environmental conditions most critical to the organism.

The movement of information can be imagined as a flow variable, running from a region of higher concentration to lower (Schneck, 1990). Thus, information, like heat or light, can have a capacity and resistance that store information and limit its flow.

The form in which information is stored in living systems is in the state of order of the various levels of biological organization, from subcellular to ecological. Beginning with the genetic code, moving on to the structures of
proteins, the specialization of the many types of cells in the body, and the intricate natural balances of species, there is order in living systems (May, 2006; Schneider, 2006). This order represents information about which we are still learning; it is this knowledge that the biological engineer must know in order to produce a successful design.

As we have previously seen (see Sections 2.5, 2.6, and 3.11), the maintenance of an ordered state requires energy expenditure, energy extracted from the environment. Entropy has previously been introduced (Section 2.5) as a concept of disorder. Thus, information storage (as an ordered state) and thermodynamic entropy (as a measure of disorder) are somehow related inversely. Shannon’s definition of information is (Shannon and Weaver, 1949; Gatlin, 1972; Loewenstein, 1999):

\[
I = \sum p_i \log_n p_i
\]  

(4.6.1)

where \( I \) is information content, \( p_i \) is probability of a particular occurrence, and \( \log_n \) is the logarithm using a meaningful base \( n \). The logarithmic base is usually chosen to be 2; in this way the unit of information is in bits (standing for binary digits), common in digital computers. As Igor Aleksander has said: The amount of information \( I \) depends on the surprise that the message holds. This is because the mathematical way of expressing surprise is as a probability \( p \); the less probable an event is, the more surprising it is and the more information it conveys.

Thermodynamic entropy of this ordered system is given as (Loewenstein, 1999):

\[
\text{entropy} = -(\text{Boltzmann’s constant}) (\ln 2) I
\]  

(4.6.2)

where Boltzmann’s constant has a value of \( 1.3802 \times 10^{-23} \text{ N·m/K} \), and \( \ln 2 = 0.693 \). By this means, it is possible to equate information storage (as structural order) in a biological system with energy storage in the same system. This also reminds us that, in order to maintain the order represented by a living system, energy must be extracted from the surrounding environment, and overall the universe becomes less ordered. Energy, information, and entropy (or order) are all related.

**Example 4.6.1 Information Content of Micrococcus DNA**

There are four bases (A, T, C, and G) in the DNA of the bacterium *Micrococcus phlei*. The probability of occurrence of each of these has been experimentally found to be (Gatlin, 1972):
\[
\begin{align*}
p(A) &= 0.164 \\
p(T) &= 0.162
\end{align*}
\]
p(C) = 0.337  
 p(G) = 0.337

Calculate the information content of the DNA of the DNA molecule in bits.

Solution:

From equation 4.6.1,

\[ I = \sum \text{p}_i \log_2 \text{p}_i \]

\[ I = 0.164 \log_2 0.164 + 0.162 \log_2 0.162 
+ 0.337 \log_2 0.337 + 0.337 \log_2 0.337 \]

Since

\[ \log_2 x = (\log_{10} x) / (\log_{10} 2) = 3.32 \log_{10} x \]

Then

\[ I = 3.32 (0.164 \log_{10} 0.164 + 0.162 \log_{10} 0.162 
+ 0.337 \log_{10} 0.337 + 0.337 \log_{10} 0.337) \]

\[ I = -1.910 \text{ bits} \]

Remark: The negative value indicates that information was extracted from the environment.

**Example 4.6.2 Entropy Value of Micrococcus DNA**

Calculate the entropy value for the DNA found in Example 4.6.1.

Solution:

From equation 4.6.2,

\[ \text{entropy} = - (\text{Boltzmann’s constant}) (\ln 2) I \]

\[ \text{entropy} = - \left(1.3802 \times 10^{-23} \frac{\text{N} \cdot \text{m}}{\text{K}}\right) (0.693)(-1.910) \]

\[ = 1.83 \times 10^{-3} \frac{\text{N} \cdot \text{m}}{\text{K}} \]

Remark:

The positive entropy value indicates that the maintenance of the DNA structure of *Micrococcus phlei* has caused disorder in the rest of the universe.
Applications and Predictions

1. Teachers will contribute information and students will accept information. Because the total amount of information storage for teachers and students together will increase, energy must be expended in the process.
2. Living systems without good memories will be more wasteful thermodynamically than those with good memories.
3. The growth process, where information will be stored at a very rapid rate, will be energy intensive.
4. Distinguishing useful information in the presence of noise will require repetition.
5. The formation of new DNA will increase the entropy of the organism.
6. Weaker signals will require more energy to recognize and process.
7. Environmental stress will require energy for coping; thus, less information can be absorbed during stress.
8. Organisms that adapt to their environments will develop improved information gathering means.
9. Information flows naturally from sources of high concentration to low concentration. A great expenditure of energy will be required to move information from low to high concentration.
10. Information storage and recall will give a survival advantage. Information storage without recall will be nonconsequential.
11. Students who are able to distinguish between important and unimportant information will expend less energy to learn.
4.7 Analog and Digital Signal Processing

For a giant like Newton, the calculation of π was chickenfeed, and indeed, in his Method of Fluxions and Infinite Series, he devotes only a paragraph of four lines to it, apologizing for such a triviality with a by the way in parentheses – and then gives its value to 16 decimal places.

-Petr Beckmann

Computation and signal processing can be performed in either one of two ways: analog or digital. Analog computation involves an infinite variety of signal levels, and useful information is represented by a value that occurs between high and low extremes. For instance, sound that you hear or light that you see is analog information. You can tell one sound intensity level from another because of its loudness, and there are many different loudness levels that can be discriminated. Many of the body’s sensors detect analog signals.

Digital signals occupy discrete levels. In the digital computer world, there are two discrete levels represented by the upper and lower power supply voltages. This is a binary system. When considering DNA, there are four discrete biochemical building blocks. This is a quaternary system. Information in a digital system does not occupy the vast array of possibilities present in an analog system. Instead, information is contained only in the presence or absence of the discrete levels.

Neural communication is digital. Detectable information in a neuron is represented by the presence or absence of an action potential. It’s either there or it isn’t, and so it’s a binary system.

Because many biological sensors detect analog levels, but communication of these levels often occurs by digital neural means, there must be an analog-to-digital conversion process in between (other communication, for example hormones, is entirely analog). Through analog-to-digital conversion, the sensed analog signal level is usually converted to a train of pulses the frequency of which corresponds to the input level (sometimes, as with taste, it is the pattern of action potentials that determines the type of taste – sweet, sour, bitter, salt, savory – that is detected).

Digital signal processing is relatively noise-free. Because information is recognized to exist only at discrete levels, small differences that occur because of environmental factors (some determined by basic physical fluctuations at the atomic or molecular level) are of no consequence. Thus, a signal level that is within a band of levels is only recognized as the standard level.

Analog processing, on the other hand, can be more efficient as long as it does not have to be particularly precise nor fast. A very precise analog computation requires very sensitive analog components, and there is a limit to how sensitive these can be. Because of thermal noise present nearly everywhere in biological systems, extreme precision must require sufficient time to average random fluctuations. If that time is not available, accuracy
can suffer. One means to increase analog computational efficiency is to transform the signal nonlinearly, as, for example, logarithmically. This transformation compresses a multi-order-of-magnitude signal into a relatively small range. This can be done with a few simple non-linear components the type of which are biologically common.

The same type of processing using digital means would require many digital channels and some fancy digital algorithms. Precision in the digital world translates into numbers of channels for parallel processing or speed for serial processing. Nonlinear transformations are almost entirely performed by analog rather than digital means.

Nature’s solution for signal processing is first to process incoming analog information efficiently with specialized analog devices such as eardrums, cochleas, and retinal cells (Sarpeshkar, 2006). The purposes of these steps are to reduce the amounts of data requiring conversion into digital form. For example, the inner ear splits analog sounds into frequency ranges before each range is transmitted to the brain by the auditory nerve. Image processing in the retina detects and locates edges, and this information is transmitted neurally to the brain.

Only after the amount of data has been reduced to conform to the conveyance limits of neurons are the signals converted to digital form. Myelinated neurons can transmit action potentials faster than slower unmyelinated neurons. Thus, in any given amount of time, more information can be conveyed on a myelinated neuron. Analog processing either limits the amount of information to the rate at which it can be transmitted, or additional information is lost.

The neuron itself functions as an analog processor, analog-to-digital converter, and digital signal transmitter. The input side of the neuron has a transmembrane potential that varies either with the incoming sensory stimulus or with the rate of pulses coming from other pre-synaptic neurons. Analog-to-digital conversion is accomplished by the presence of a threshold voltage above which the action potential is formed. The action potential is the digital form of the signal.

Applications and Predictions

1. Living things will use the type of processing that is most efficient for the function that is to be performed.
2. Digital communication lines are less likely to be affected by environmental influences. Thus, neuronal communication systems predominate over other types.
3. Knowing the magnitude of a stimulus is important to formulate a response. Therefore, primary receptors will most likely be analog, with the analog signal somehow being able to be sensed by the organism.
4. Simple creatures will operate almost exclusively in the analog world.
QUESTIONS
Chapter 4

4.0.1 What place does mathematics have in bioengineering methods?

4.0.2 List the engineering science courses you expect to take and give an assumed relationship to biological engineering for each of these.

4.0.3 State why the study of mathematics contributes to biological engineering understanding.

4.0.4 State why the study of engineering sciences is important for biological engineers.

4.0.5 How does a mathematical model lead to an improved biological engineering prediction?

4.0.6 What is an engineering science and how does it relate to the study of biology?

4.0.7 What are the advantages and disadvantages to applying mathematical analysis to biological systems?

4.0.8 Why is the understanding of control systems important to the study of biological systems?

4.1.1 In the equation,

\[ \text{power} = 38.2 \text{ (speed)}^2 \text{ (grade)} + 9.8 \text{ (mass) (speed)}, \]

where mass is in kg and speed in m/sec, what are the units of power, the coefficient 9.8, and the coefficient 38.2? Grade is a dimensionless fraction.

4.1.2 The quadratic formula is:

\[ x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \]

Is this an unequivocal equality, conditional equality, or replacement function?

4.1.3 In the equation 2.5.2 defining enthalpy, the following variables and dimensions are evident:

heat (H)
work (F·L)
pressure (F / L²)
volume (L³)

How many independent p, terms can be found? Find one of these.

4.1.4 Describe the process of iteration. Why is iteration only a conditional
equality? Give examples of where iteration can be applied.

4.1.5 Add to the list of Applications and Predictions.

4.2.1 What do you think you would have to know in order to know exactly
the state of some biological system? Is it possible to know these
things with certainty?

4.2.2 What sources of variation are there in a measurement of a biological
system that are not directly attributable to the system?

4.2.3 Give examples of biological measurements that are likely to conform
to each of the probability distribution functions given in Figure 4.2.1.

4.2.4 Give examples of things that are interesting because they are almost
predictable, but not quite.

4.2.5 For the following sets of data, determine the best fit lines by least
squares. Next, graph the data and judge how well the least squares
lines represent the function describing the data.

<table>
<thead>
<tr>
<th>x</th>
<th>y₁</th>
<th>y₂</th>
<th>y₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.0</td>
<td>8.04</td>
<td>9.14</td>
<td>7.46</td>
</tr>
<tr>
<td>8.0</td>
<td>6.95</td>
<td>8.14</td>
<td>6.77</td>
</tr>
<tr>
<td>13.0</td>
<td>7.58</td>
<td>8.74</td>
<td>12.74</td>
</tr>
<tr>
<td>9.0</td>
<td>8.81</td>
<td>8.77</td>
<td>7.11</td>
</tr>
<tr>
<td>11.0</td>
<td>8.33</td>
<td>9.26</td>
<td>7.81</td>
</tr>
<tr>
<td>14.0</td>
<td>9.96</td>
<td>8.10</td>
<td>8.84</td>
</tr>
<tr>
<td>6.0</td>
<td>7.24</td>
<td>6.13</td>
<td>6.08</td>
</tr>
<tr>
<td>4.0</td>
<td>4.26</td>
<td>3.10</td>
<td>5.39</td>
</tr>
<tr>
<td>12.0</td>
<td>10.84</td>
<td>9.13</td>
<td>8.15</td>
</tr>
<tr>
<td>7.0</td>
<td>4.82</td>
<td>7.26</td>
<td>6.42</td>
</tr>
<tr>
<td>5.0</td>
<td>5.68</td>
<td>4.74</td>
<td>5.73</td>
</tr>
</tbody>
</table>

4.2.6 Is chaotic data random? Why or why not?
4.2.7 Should the strength of a disinfectant be based on the value that is lethal for the average microbe? Why or why not?

4.2.8 Where do statistical methods fit into the scientific method?

4.2.9 Describe a very large experiment. How can this experimental plan be modified to cost less to conduct?

4.2.10 Why should treatment subgroups be carefully considered when analyzing them for statistical significance?

4.2.11 Muscle fatigue can be measured by means of the frequency spectrum, related to standard deviation, of electrical signals recorded from the muscle. The following is a list of values (in arbitrary units) for the myoelectric signal during isokinetic knee extension using the rectus femoris muscle. Compute the standard deviation.

\[
\begin{align*}
0.35 & \quad 0.15 & \quad 0.70 & \quad -0.65 & \quad 0.28 & \quad -0.10 \\
-0.25 & \quad -0.24 & \quad -0.80 & \quad 0.16 & \quad -0.35 & \quad -0.03 \\
-0.10 & \quad 0.15 & \quad 0.04 & \quad 0.00 & \quad 0.04 & \quad 0.00 \\
-0.10 & \quad -0.13 & \quad 0.50 & \quad 0.25 & \quad 0.08 & \quad 0.06 \\
0.08 & \quad -0.23 & \quad 0.22 & \quad 0.01 & \quad 0.08 & \quad 0.01
\end{align*}
\]

4.2.12 In developing a countertop spray cleaner effective against Salmonella bacteria, candidate antibiotics are tested on agarose plates inoculated with Salmonella bacteria. Single drops (0.1 ml) of 1 M antibiotic solutions are introduced to each plate, and the zones of Salmonella inhibition are measured after a standard incubation time. The standard for the experiment is a 70% solution of ethanol in water. Diameters (in cm) for each of 16 replications for each antibiotic are given below. Which antibiotic do you recommend?

<table>
<thead>
<tr>
<th>Sodium chlorite</th>
<th>2-Benzyl 4-chlorophenol</th>
<th>Dehydrobiethylanine</th>
<th>70% Ethanol</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.85</td>
<td>0.95</td>
<td>1.47</td>
<td>1.02</td>
</tr>
<tr>
<td>1.58</td>
<td>0.90</td>
<td>1.71</td>
<td>0.80</td>
</tr>
<tr>
<td>1.50</td>
<td>1.06</td>
<td>1.25</td>
<td>0.94</td>
</tr>
<tr>
<td>1.19</td>
<td>0.86</td>
<td>2.15</td>
<td>0.88</td>
</tr>
<tr>
<td>1.87</td>
<td>1.32</td>
<td>2.10</td>
<td>0.85</td>
</tr>
<tr>
<td>1.46</td>
<td>1.20</td>
<td>1.57</td>
<td>0.90</td>
</tr>
<tr>
<td>1.73</td>
<td>1.08</td>
<td>2.21</td>
<td>0.89</td>
</tr>
<tr>
<td>1.58</td>
<td>0.99</td>
<td>2.03</td>
<td>0.76</td>
</tr>
<tr>
<td>1.50</td>
<td>1.24</td>
<td>1.63</td>
<td>0.99</td>
</tr>
<tr>
<td>1.78</td>
<td>0.93</td>
<td>2.74</td>
<td>0.93</td>
</tr>
<tr>
<td>1.76</td>
<td>1.16</td>
<td>1.92</td>
<td>0.85</td>
</tr>
<tr>
<td>1.36</td>
<td>1.18</td>
<td>1.87</td>
<td>0.91</td>
</tr>
<tr>
<td>0.76</td>
<td>1.07</td>
<td>2.40</td>
<td>0.75</td>
</tr>
<tr>
<td>1.02</td>
<td>1.04</td>
<td>0.81</td>
<td>0.81</td>
</tr>
<tr>
<td>1.45</td>
<td>1.06</td>
<td>1.58</td>
<td>0.77</td>
</tr>
<tr>
<td>1.30</td>
<td>1.27</td>
<td>1.25</td>
<td>0.85</td>
</tr>
</tbody>
</table>
4.2.13 What are the dangers from drawing conclusions from small sample sizes?

4.2.14 How is grouping according to sources of variation related to individualized medicine?

4.2.15 Add to the list of Applications and Predictions.

4.3.1 How are position, velocity, and acceleration related? Write a differential equation relating these three variables.

4.3.2 Why are few biological responses proportional to \( e^{t/\tau} \)?

4.3.3 Discuss the advantages to a biological system of the exponential response to a sudden change (step input) as given in Figure 4.3.1.

4.3.4 Where, in biological systems, are periodic responses found? Are these the result of intrinsic periodicity or forced by environmental oscillations?

4.3.5 If volume flow rate is \( V_o \sin \omega t \), where \( t = \) time, \( \omega = \) frequency, and \( V_o = \) magnitude, what is \( \frac{dV}{dt} \)? What is the amount of work, as given in equation 4.3.10?

4.3.6 Add to the list of Applications and Predictions.

4.4.1 List as many bodily sensors as you can. Compare these to sensors in plants. Are they similar?

4.4.2 If you had to replace a natural sensor with one made by humans, how would you do so?

4.4.3 List actuators in addition to muscles. What are the results of their actions?

4.4.4 Would you guess the efficiency of fast twitch fibers to be greater or less than that of slow twitch fibers?

4.4.5 Sometimes cardiac insufficiency is corrected surgically by grafting a piece of skeletal muscle to the heart ventricle. What would you think would be the difficulty with using skeletal muscle in this way? Could skeletal muscle substitute for smooth muscle?
4.4.6 Speculate on why two communication systems are necessary in the body. Why can plants do fine with one?

4.4.7 If resistance, capacity, and inertia of neurons slows and degrades the signal, what mechanisms have evolved to counteract untoward effects?

4.4.8 Distinguish between the sympathetic and parasympathetic nervous systems. Which dominates at rest and which during exercise?

4.4.9 Describe the process forming action potentials. How could you use this knowledge to produce neural prostheses?

4.4.10 Describe the time and space scales of synaptic neural transmission. What challenges do these scales present?

4.4.11 What is meant by feedback and feedforward?

4.4.12 Make a list of examples of open-loop control in biology.

4.4.13 Make a list of apparent feedforward control in biology.

4.4.14 Would you expect adaptive control to be more or less widely present in living things? Why?

4.4.15 Compare neurotransmission times and distances with other biological metrics.

4.4.16 Develop a fuzzy control algorithm to go from your home to a favorite store.

4.4.17 Describe the neural transmission of a signal down one neuron, from that neuron to the next, and down the second neuron.

4.4.18 How would you go about determining events occurring in the interneuronal gap?

4.4.19 Add to the list of Applications and Predictions.

4.5.1 How would optimization be incorporated into an engineering design involving living things?

4.5.2 Compare expected performance of a biological organism to a narrow versus broad optimum.

4.5.3 Add to the list of Applications and Predictions.
4.6.1 What is information?

4.6.2 How is information related to order in biological systems? Are the two synonymous?

4.6.3 Add to the list of Applications and Predictions.

4.7.1 Explain differences between analog and digital signals. Give common examples of each.

4.7.2 Picture a bioreactor. What processes of the bioreactor are treated as analog and which as digital?

4.7.3 What parts of a neuron are analog and what parts are digital?

4.7.4 Add to the list of Applications and Predictions.