CHAPTER 2

PERCEPTION AS A CHOICE AMONG ALTERNATIVES

PERCEPTION: UNDEFINED

In order to define a word, we try to express it in terms of other words whose meanings are, perhaps more familiar. For example, Webster's definition of a pen is "Any instrument for writing with ink ..." This process is fine as far as it goes, but certain words denote ideas that seem to defy expression in simpler form. Of this genre is *perception*, which Webster's approaches valiantly as "The act of perceiving; apprehension with the mind or the senses; ..." However, comprehending *mental apprehension* may be no simpler than comprehending *perception* itself. Coren and Ward (1989) wisely opt in favor of an operational definition: "Someone who is interested in the study of perception is interested in our conscious experience of objects and object relationships." So if we understand what is meant by consciousness we can infer the meaning of *perception*. The difficulty in definition does not in any way deter us from using the word. The noun, *perception* seems to be growing in popularity, particularly among journalists. Have you noticed how frequently we hear of "a growing perception of ..." or "my perception of the situation is that ..."? I must admit, that despite the plethora of books on the subject, and the unquestioned erudition of their authors, I often introduce *perception* to my students using exactly the same example as my grade 2 teacher, Jenny Snider, used for me when I was seven years old.

"My friend asked me the time," she told me, "and I looked at my watch."

"So what is the time?" reiterated her friend.

"Oh!" replied my teacher. "I looked at my watch, but I didn't perceive the time!"

Although she had *mentally apprehended* her watch, and was definitely *conscious* of having done so, she, nonetheless, had not *perceived* the time.

So, definitions set aside, what is the salient feature of *perception*? What characterizes this form of human activity? Does it lie uniquely within the province of the sciences of physiology and psychology? Is it shared by animals? Can we build automata, robots, that will also partake in this activity?

I am going to argue that the crux of the process of perception was appreciated by the ancients; that perception can be treated as legitimately by physics as by psychology; that animals perceive; and that if we construct our robots with sensory systems embracing both active and passive components (to be discussed later), these robots will, in principle, be able to perceive.

Let's start with the ancients – at least with the ancient Romans. I was miffed, a few years ago, when I learned from the Department of Classics at the University of Toronto that the Roman-in-the-street would not likely have said "Percipio" if he or she meant "I perceive" (I had dutifully ferreted this out of a Latin dictionary), but rather "intellego." Wherefore "intellego"? Well, etymologically the word is made up of two simpler Latin words: *inter* meaning *between*, and *lego* meaning *I choose* or *gather*. *To perceive* was, then, *to choose between* alternatives.

I like it. *To choose between*. I like it because it codifies what I believe is a fundamental concept underlying and unifying perceptual phenomena: making a choice among alternatives. For example, if one perceives that the color is blue, they have chosen blue from among other possible colors: red, orange, green, ... If Mrs. Snider had perceived that it was ten o'clock, then she had selected ten from among the possibilities one, two, ..., twelve. When she looked at her watch without making this selection, she, of course, failed to perceive the time. When the TV anchorperson remarked that there was a perception among meteorologists that a global greenhouse (warming) effect was imminent, the implication was that "greenhouse" was selected from among at least the possibilities "greenhouse," "icehouse" or "no effect."

Perception involves a selection from among alternatives, but this property of perception does not, in itself, involve the idea of consciousness. For example, a simple Meccano model might select long metal strips from among an assortment of long and short strips, but we would be hard pressed to attribute to it any form of consciousness. We shall do the best we can with the concept of consciousness as the theory unfolds.

Perception, as an *intellego*-activity, implies something important. It implies that the *percept* (which we shall take to mean *that which is perceived*) must always be found among a set of alternatives contained within the memory of the perceiving system. Thus, for example, if I perceive that an apple is large, it is because I retain in my memory a record of apples of various sizes, and I can make the selection of "large" from among them.[1] Since perception is dependent on the memories of the perceiving system, it follows that if different perceiving systems have different memories, they will perceive differently. Within the sphere of the very simple stimuli that we shall be considering it means, for example, that a sensory system for perceiving the brightness of light can perceive only those levels of brightness that are contained within its "memory." It is certainly tempting to ask: Well, how did one perceive these aboriginal brightnesses and temperatures? In this matter, I shall again beg the reader's patience. We shall return to this question in due course.

As a consequence of the above, we can say that perception is *relative* to the memory of the perceiving system. Perception is, therefore, *relative-intellego* in nature: *intellego* insofar as it *chooses* among alternatives, and *relative* in that it depends on the state of the perceiving system.

We should also observe that the *intellego* concept allows for approximate perception. That is, one might select not just one member of a set of alternatives but several members. For example, one might perceive that a color is either orange or red (but not green or blue); and I may perceive that your age lies between 35 and 40 (but you are definitely not a teenager). Perception can, therefore, denote a reduction in the number of alternatives without arriving at a unique remaining possibility.

INFORMATION: DEFINED

We are all familiar with the term "information" in the semantic or verbal sense. For example: The newspaper contains a lot of information.

Or: I lack the information needed to reach a destination.

However, the term "information" is usually assigned a meaning in the generic sense, which removes it from any particular context. Following the seminal work by Shannon (1948) and Wiener (1948), *information* is defined as a reduction in uncertainty. Suppose that an event may occur in one of many ways; that is, it has many possible outcomes. We might say that we are *uncertain* about the outcome. For example, if a coin is tossed, it can fall heads or tails, so that there is some momentary uncertainty about the outcome. After the outcome is known, the uncertainty vanishes, or is resolved, and the perceiver[2] of the event receives a quantity of information equal to the vanished uncertainty. *Information theory* provides a way of quantifying the initial uncertainty and, therefore, quantifying the information received. The uncertainty preceding the occurrence of an event is usually termed *entropy*, so that the quantity of information received is equal to the reduction in entropy. One must be careful not to confuse this type of entropy, which is information theoretical entropy, with *physical* or *thermodynamic entropy*, a quantity used extensively in physics and chemistry.[3] While there is a relationship between the two entropies, which we shall explore later on, the entropy used in this book will generally be of the information theoretical type. The symbol H will be used to designate entropy.

In the coin-tossing example, there were only two possible outcomes but, in general, an event *with discrete outcomes* can have x possible outcomes, where x is a positive integer.[4] Suppose that the probability of the i^{th} possible outcome is p_i. Then entropy, H, is defined by the equation

$$H = -\sum_{i=1}^{x} p_i \log p_i \,. \tag{2.1}$$

That is, H is composed of a weighted sum of the logarithms of the probabilities of the outcomes. The base of the logarithms used is arbitrary. We shall generally use natural logarithms (ln), logarithms to the base e, because natural logs are simpler to deal with in many mathematical operations. The "unit" of H will be the *natural unit*. When logarithms to the base 2 are used, the "unit" of H becomes the *bit*. It is a simple matter to convert from one unit to another:[5]

$$\text{bits} = \text{natural units} / \ln 2 . \tag{2.2}$$

Therefore, when entropy, H, is reduced totally, that is, when one of the x outcomes is perceived to occur, the amount of information, $\mathscr{I}$, that is, received is given simply by

$$\mathscr{I} = H . \tag{2.3}$$

To take a simple example, if the event is the selection of one card from a shuffled deck of 52, the probabilitiesp_i are all equal to $1/52$ so that

$$\begin{aligned} H &= -(1/52)\log(1/52)\ \ldots\ -(1/52)\log(1/52) \\ &= -52/52\log(1/52) = \log 52 . \end{aligned}$$

If we use natural logs, $H = \ln 52 = 3.95$ natural units $= 3.95/\ln 2$ or 5.70 bits of information. From this example we can infer the general theorem that the entropy of an event with x outcomes, all of which are equally probable, is equal to $\log x$.

It is, of course, possible that there is some interfering factor that prevents the perceiver from determining the outcome of an event with complete certainty and, therefore, the number of possible outcomes may not be reduced to one, single possibility. That is, some residual uncertainty may remain. For example, the event may be hearing a spoken numeral between one and nine. Suppose that each of the 10 possible numerals are equally probable so that (analogous to the playing card example above) the entropy equals log 10. However, the spoken words "five" and "nine" are easily confused, so it may be that even after the event has occurred (the spoken numeral is heard) there is still some residual uncertainty about whether it was "five" or "nine". The entropy has been reduced from log 10 to log 2, but not to $\log 1 = 0$. In this case, $\mathscr{I} < H$: the transmitted information is less than the total original entropy (Brillouin, 1964).

Notice also the dependence of information receipt, $\mathscr{I}$, on the set of values of the *a priori* probabilities, p_i. Different values for p_i will give rise to quite different values for $\mathscr{I}$. For example, suppose I toss a coin that you think is a fair coin with equal probabilities for head and tail, but that I know is biased. I know, from the manner in which the coin was constructed, that it will fall heads on $2/3$ of the tosses and tails on only $1/3$ of the tosses. Then

$$H_{\text{you}} = -(1/2)\log(1/2) - (1/2)\log(1/2) = 1 \text{ bit},$$

while

$$H_{\text{me}} = -(2/3)\log(2/3) - (1/3)\log(1/3) = 0.92 \text{ bit}.$$

I have received less information than you have although we both perceived the outcome of the same event. The reason why I received *less* information is not as important as the fact that we received *different amounts*. Our respective previous life experiences have determined our value for the *a priori* probabilities and, hence, for the received information. Received information is, in this respect, not an absolute quantity.

The reader may already have appreciated the parallelism between the concept of *relative intellego perception* and that of information by entropy reduction, which is the subject of the next section.

INFORMATION AND PERCEPTION

The *relative intellego* nature of perception, as we have seen, implies that perception involves a selection of one or several choices from among many, and that the choices available were dependent upon the history and memory of the perceiving system. We have also learned that a measure of information is obtained by making a selection of one or several possible outcomes of an event from among many possible outcomes, and that the quantity of information is calculated from the *a priori* probabilities, which are, in turn, dependent upon the history and memory of the recipient of the information. From this parallelism it may be seen that information theory provides a natural means to measure the "quantity" or "amount" of perception. Let's take a few examples. On a macroscopic scale, if I perceive that you are wearing your brown jacket (from your wardrobe of two, equally-worn jackets) my percept is "worth" $\log_2 2$, or one bit of information, while if I perceive that you are wearing your green tie from your holdings of 8, equally-worn ties, my percept is "worth" $\log_2 8$, or three bits of information (the selections are assumed independent since you are an academic and are, therefore, unlikely to match tie with jacket). In order to apply this idea at the microscopic scale of perception at this stage in our development of information theory, let us pretend that for some hypothetical creature, light intensity is perceived only in discrete levels: intensity level 1 (lowest level), intensity level 2 (a little more intense), ..., intensity level x (most intense possible level). For a particular percept, the probabilities of the respective light intensity levels are known to be p_1, p_2, ..., p_x. Then the quantity of information associated with the percept of exactly one of these intensity levels is equal to $-p_1 \log p_1 - p_2 \log p_2 \; ... -p_x \log p_x$. Information theory is a natural means of quantifying a percept.

We should note in passing that one can also quantify an approximate percept. Continuing the example of light intensities, let us suppose that there were originally 10 identifiable light intensity categories, so that the uncertainty before the act of perception is given by

$$H_{\text{before}} = -\sum_{i=1}^{10} p_i \log p_i \; .$$

If, just for convenience of discussion, all the p_i were equal, then

$$H_{\text{before}} = \log 10 \; .$$

Suppose, now, that the process of perception serves to reduce the number of intensity categories to three rather than to a single category, and that the three are equally probable. Then

$$H_{\text{after}} = \log 3 \; .$$

That is, information received, $\mathscr{I}$, is given by

$$\begin{aligned} \mathscr{I} &= H_{\text{before}} - H_{\text{after}} \\ &= \log 10 - \log 3 \; . \end{aligned} \tag{2.4}$$

In all cases, information received is relative to the number of intensity categories and their respective *a priori* probabilities.

The above example does not indicate exactly the manner in which we shall be carrying out calculations later on, but it does introduce the general idea.

We might observe, at this point, that information theory also provides a natural means of quantifying what might be called "quantum perception." In quantum physics, the anticipation of a measurement is associated with a wavefunction, which is a mathematical construct representing a number of possible outcomes to the measurement. Each possible outcome can be assigned a probability. When the measurement is actually made, the wavefunction "collapses," leaving only a single possibility. In this respect, it can be seen that one could assign an information content to the quantum measure.

Suppose that a quantum system consists of a *mixture* of two possible states, which we can designate as $|\psi_1\rangle$ and $|\psi_2\rangle$. A measurement performed on the system will disclose one or other of the two states. The probability of $|\psi_1\rangle$ is p_1 and the probability of $|\psi_2\rangle$ is p_2. The information content of this percept is then

$$\mathscr{I} = -p_1 \log p_1 - p_2 \log p_2 \; . \tag{2.5}$$

A *superposition* of quantum states (vis-à-vis a mixture) will give rise to a more complicated set of probabilities, but the principle is the same. Quantum mechanical perception (usually called "observation") can be seen to be an instance of what we have called *intellego perception*, the selection of one from several possibilities.

THE GIST OF THE ENTROPIC THEORY OF PERCEPTION

Hitherto, we have made no biological commitment. We have calculated H-functions for various simple perceptual situations, but we have not yet related the entropy, H, to any biological function. In the example of light intensity explored above, we calculated that $\log 10 - \log 3$ units of information were gained in the process of perception; but we have not specified any significance of this information to the perceiving organism. Indeed, why *should* there be any significance? I shall argue in various ways throughout this book that there is a great deal of biological significance.

The fundamental equation to be put forward is that

$$F = kH \,, \tag{2.6}$$

where k is a constant, greater than zero, and F is a *perceptual variable*. An example of F, albeit not one of which I am overly fond, is the subjective magnitude of a stimulus. Drawing again upon the example of perception of light intensities, the physical intensity of a steady light source could be measured in watts (leaving aside more complex photometric units). From the physical intensity we can calculate, in the manner shown above, the value of the entropy, H. Multiplying H by a constant, we obtain F, which is the subjective magnitude of the light stimulus, or simply the *brightness* of the light. The value of the constant, k, is determined by the scale on which the investigator wishes to measure brightness.

Just to render the idea a little more concrete, let us take a numerical example. Let's continue to suppose that some organism can perceive only discrete intensities of light, rather than a continuum of intensities as we human beings can. Suppose that after a brief exposure to a light stimulus, the organism can determine that the *intensity* of the light, that is, the power of the light source is one of 30 watts, 40 watts, or 50 watts, with equal probabilities of $1/3$ assigned to each. Its perceptual entropy or uncertainty, as obtained from Equation (2.1), is equal to $\log_2 3 = 1.58$ bits. Suppose that a scale of brightness is set up with $k = 1$. Since $F = kH$ from Equation (2.6), the brightness of the source would be equal to $1 \times \log_2 3 = 1.58$ scale units. Undoubtedly, this calculation will raise a number of questions in the mind of the thoughtful reader,[6] but let's give the ideas time to unfold naturally.

While this example does capture the gist of the idea that will be developed, the restriction of light sensitivity to discrete values is artificial, and the reader should resist comparing the result with the well-known laws of human sensation that apply to continuously variable stimuli.

A second example of a perceptual variable, F, is the impulse rate, or the rate at which action potentials propagate, in a primary sensory afferent neuron. Imagine a single light detector dissected out of the retina of an animal, still attached to its sensory neuron. For example, a single omatidium from the eye of the *limulus* crab was investigated some years ago by Hartline and Graham (1932). The investigator can then stimulate the photoreceptor *in vivo*, and measure the frequency of action potentials (electrical spikes) that propagate down the attached primary sensory neuron (nerve fiber). From the physical intensity of the stimulating light we shall be able to calculate the value of the entropy, H (although we have not yet described explicitly how this can be done), and equate this value to a constant, k, multiplied by the impulse frequency, F. So, for example, if $H = 3$ bits and $k = 10$ [impulses $\cdot$ $s^{-1} bit^{-1}$], the impulse frequency, $F = kH = 30$ impulses or action potentials per second. The key is, of course, to be able to evaluate H for continuously distributed light signals.

We have now carried the idea of the previous section one step further. In the previous section, we showed that a percept could be quantified using information theory. In this section we can begin to understand the psychophysical (e.g. brightness) and physiological (e.g. impulse rate) impact of the quantified percept on the biological organism. Brain and neuron respond "in proportion to" the entropy of the stimulus.

Equation (2.6), $F = kH$, is a conjecture which we shall be exploring in some depth. Its origins and justification will have to await the final chapters of the book, but where does this equation lead? We shall see that it leads toward the unification of the "laws" of sensation. That is, it will permit us to derive from a single equation many of the empirical laws of sensation, such as Fechner's and Weber's laws. It will also permit us to make a number of predictions – some of them rather unexpected – about the behaviour of the senses. It will provide some critical links between biology and physics, and may even provide some insights into the form and meaning of some of the laws of physics. Parenthetically, it will also have something to say about practical matters, such as the construction of cochlear implants for the hearing impaired. However, most important in my opinion, $F = kH$ codifies the manner in which the brain and central nervous system apprehend the "external world." In the generally held world view (recall D. Marr's remarks), the brain serves essentially to encode and discriminate "that which is out there." $F = kH$, however, casts doubt on that rather simple view. F is, in fact, the sensory language of the brain; it encodes the electrical signals that are carried from the sensory receptors to the brain. H, however, is not just a mathematical transcription of elements of the external world. If it were, we would, indeed, have a Marr world view, because the brain-picture, F, would just be a map of the external-world-picture, H. The mapping process external $\rightarrow$ *brain* would represent a completely objective process. However, H is an *uncertainty*; it reflects not what the world *is*, but rather what the sensory receptors *think* the world is. $F = kH$ implies that we only perceive those aspects of the "external" world that do not conform to our expectation, and, hence, of which we are uncertain. Moreover, and most startling, when our uncertainty vanishes ($H = 0$), so do our perceptions ($F = 0$).

I recognize that a lot of ideas have been introduced in this chapter, and rather succinctly. Before exploring these ideas further, I recommend that we take a one-chapter digression to discuss certain aspects of the experimental investigations of sensation and perception. We shall pick up the thread of $F = kH$ again in later chapters. The reader for whom *information theoretical entropy* is a new concept might like to peruse a primer on the subject before proceeding, or to read about and beyond the subject from an expert like J. R. Pierce (1980).

NOTES

1. Let us ignore the case where I observe an "apple" that is larger than any apple I have ever seen before, and I must decide if it is really an exceptional specimen of the apple species or some other kind of fruit. Such considerations would take us too far afield.

2. I have used the word "perceiver" without yet justifying it, but the reader may realize immediately why I have done so.

3. I like Peters' term, *intropy*, for *in*formation *t*heoretical ent*ropy* (Peters, 1975).

4. We shall later consider events whose outcomes lie on a continuum, such as the intensity of a sound; but for the moment we shall deal only with those events that have discrete outcomes, such as the roll of a die, whose outcome may be simply one of 1, 2, 3, 4, 5 or 6.

5. To convert H from units calculated in logs base a to units calculated in logs base b, we use the algebraic identity

$$\log_a y = \log_b y / \log_b a .$$

[Think of the mnemonic:]

$$\frac{y}{a} = \frac{y}{b} / \frac{a}{b} .$$

Therefore, $\log_2 y = \log_e y \,/\, \log_e 2$.

6. In the matter of the hypothetical organism that can perceive only discrete intensities of light ...

$$F = kH$$

with $k = 1$ becomes

$$F(\text{brightness}) = H(\text{entropy or uncertainty}).$$

Therefore, after a brief exposure of t_0 seconds to a light stimulus, the organism's photoreception system retains the uncertainty of three possible discrete light intensities: 30, 40 or 50 watts, each intensity associated with a probability of $1/3$. Therefore,

$$F = H = \log_2 3 = 1.58 \text{ "bits" of brightness.}$$

After a little thought, however, the reader may conclude that this process is patently wrong. After all, it is possible that a second stimulus will be less intense, and after the t_0 seconds of exposure, the organism's photoreception system will again experience the uncertainty of three discrete light intensities: this time 10, 20 and 30 watts. Again we would have

$$F = H = \log_2 3 = 1.58 \text{ "bits" of brightness.}$$

That is, the less intense light stimulus would appear equally as bright as the more intense stimulus, which is counterintuitive. Indeed, when we turn again in Chapter 16, to the hypothetical case of an organism which can perceive only discrete intensities, we tolerate the consequence within the entropy theory that all stimuli appear equally bright. However, at the present stage in our study, we shall attempt to avoid this strange result by not prodding too deeply.

Resolution of this paradox, at least superficially, is found in what has been called the "Fullerton-Cattell law," an empirical law cited by G.S. Fullerton and J. McK. Cattell in 1892: "The error of observation tends to increase as the square root of the magnitude, the increase being subject to variation whose amount and cause must be determined for each special case." The law was generalized by Guilford (1932) to read (using our symbols)

$$\Delta I = kI^n ,$$

where I is the intensity of the light stimulus, and ΔI is the magnitude of the uncertainty after t_0 seconds. k and n are constants that are greater than zero. That is, the range of observable stimulus intensities increases as the nth power of the mean intensity.

As applied to our problem, suppose that the sensory receptor takes several samples of its stimulus. We shall apply Guilford's law to mean that the range of light intensities, ΔI, recorded by the receptor for the smaller stimulus is smaller than the range of light intensities recorded for the larger stimulus. Since the larger stimulus left the photoreceptor system with an uncertainty of three light intensities, the smaller stimulus would leave the receptor with an uncertainty of two light intensities, say 10 and 20 watts, or even with the certain result of a single intensity. Therefore, for the less intense light stimulus,

$$F = H = \log_2 2 = 1 \text{ " bit" of brightness,}$$

or

$$F = H = \log 1 = 0 \text{ [imperceptible stimulus].}$$

That is, the greater stimulus appeared brighter than the smaller stimulus, not directly because it was more intense, but because it engendered greater entropy or uncertainty at the level of the photoreceptor system.

Applying Guilford's law inappropriately, as we have done above, to an isolated sensory receptor is a teaching device only. This law will be derived later, in effect, when we come to treat the Weber fraction theoretically.

REFERENCES

Brillouin, L. 1964. Science, *Uncertainty and Information*. Academic Press, New York.

Coren, S. and Ward, L.M. 1989. *Sensation and Perception*, 3rd Edition Harcourt, Brace, Jovanovich, San Diego.

Guilford, J.P. 1932. A generalized psychophysical law. *Psychological Review* **39**, 73-85.

Hartline, H.K. and Graham, C.H. 1932. Nerve impulses from single receptors in the eye. *Journal of Cellular and Comparative Physiology*, **1**, 277-295.

Peters, J. 1975. Entropy and Information: Conformities and Controversies. In: *Entropy and Information in Science and Philosophy*, L Kubá and J. Zeman, Editors, Elsevier, Amsterdam.

Pierce, J.R. 1980. *An Introduction to Information Theory: Symbols, Signals and Noise*. Second edition. Dover, New York.

Shannon, C.E. 1948. A mathematical theory of communication. *Bell System Technical Journal* **27**, 379-423.

Note: *All the papers of Shannon referenced in this book can be found conveniently bound together in*:

Slepian, D., editor 1974. *Key Papers in the Development of Information Theory*. IEEE Press, New York.

Wiener, N. 1948. *Cybernetics*. The Technology Press M.I.T.