# 1 Psychophysics I: Introduction and Thresholds





PORTIA: *That light we see is burning in my hall*
*How far that little candle throws its beams.*
*So shines a good deed in a naughty world.*
NERISSA: *When the moon shone we did not see the candle.*
PORTIA: *So doth the greater glory dim the less.*
*A substitute shines brightly as a king*
*Unto the king be by and then his state*
*Empties itself as doth an inland brook*
*Into the main of waters.*
*The Merchant of Venice*, Act V, Scene 1.

## 1.1 Introduction and Terminology

Psychophysics is the study of the relationship between energy in the environment and the response of the senses to that energy. This idea is exactly parallel to the concerns of sensory evaluation – how we can measure peoples' responses to foods or other consumer products. So in many ways, sensory evaluation methods draw heavily from their historical precedents

*Quantitative Sensory Analysis: Psychophysics, Models and Intelligent Design*, First Edition. Harry T. Lawless.

in psychophysics. In this chapter we will begin to look at various psychophysical methods and theories. The methods have close resemblance to many of the procedures now used in sensory testing of products.

Psychophysics was a term coined by the scientist and philosopher Gustav Theodor Fechner. The critical event in the birth of this branch of psychology was the publication by Fechner in 1860 of a little book, *Elemente der Psychophysik*, that described all the psychophysical methods that had been used in studying the physiology and limits of sensory function (Stevens, 1971). Psychophysical methods can be roughly classified into four categories having to do with:

- absolute thresholds,
- difference thresholds,
- scaling sensation above threshold, and
- tradeoff relationships.

A variety of methods have been used to assess absolute thresholds. An absolute threshold is the minimum energy that is detectable by the observer. These methods are a major focus of this chapter. Difference thresholds are the minimum amount of change in energy that are necessary for an observer to detect that something has changed. Scaling methods encompass a variety of techniques used to directly quantify the input–output functions (of energy into sensations/responses), usually for the dynamic properties of a sensory system above the absolute threshold. Methods of adjustment give control of the stimulus to the observer, rather than to the experimenter. They are most often used to measure tradeoff functions. An example would be the tradeoff between the duration of a brief flash of light and its photometric intensity (its light energy). The observer adjusts one or the other of the two variables to produce a constant sensation intensity. Thus, the tradeoff function tells us about the ability of the eyes to integrate the energy of very brief stimuli over time. Similar tradeoff functions can be studied for the ability of the auditory system to integrate the duration and sound pressure of a very brief tone in order to product a sensation of constant loudness.

Parallels to sensory evaluation are obvious. Flavor chemists measure absolute thresholds to determine the biological potency of a particular sweetener or the potential impact of an aromatic flavor compound. Note that the threshold in this application becomes an inverse measure of the biological activity of the stimulus – the lower the threshold, the more potent the substance. In everyday sensory evaluation, difference testing is extremely common. Small changes may be made to a product, for example due to an ingredient reduction, cost savings, nutritional improvement, a packaging change, and so forth. The producer usually wants to know whether such a change is detectable or not. Scaling is the application of numbers to reflect the perceived intensity of a sensation, and is then related to the stimulus or product causing that sensation. Scaling is an integral part of descriptive analysis methods. Descriptive analysis scales are based on a psychophysical model and the assumption that panelists can track changes in the product and respond in a quantitative fashion accordingly.

The differences between psychophysics and sensory evaluation are primarily in the focus. Psychophysics focuses on the response of the observer to carefully controlled and systematically varied stimuli. Sensory evaluation also generates responses, but the goal is to learn something about the product under study. Psychophysical stimuli tend to be simple (lights or tones or salt solutions) and usually the stimulus is varied in only one physical attribute at a time (such as molar concentration of salt in a taste perception study). Often, the resulting change is also unidimensional (such as salty taste). Products, of course, are multidimensional, and changing

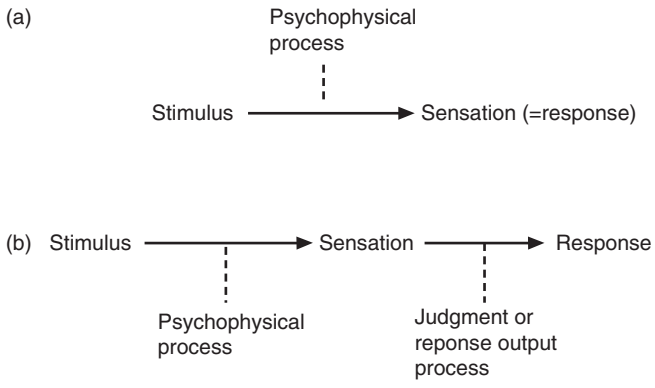


**Figure 1.1** Stimulus response models in psychophysics. A classical model has it that the stimulus causes a sensation which is directly translated into an accurate response. A more modern model recognizes that there can be decision processes and human judgment involved in translating the sensation in a response (a data point) and, thus, there are at least two stages and two processes to study.

one ingredient or aspect is bound to have multiple sensory consequences, some of which are hard to predict. Thus, the responses of a descriptive analysis panel, for example, often include multiple attributes. However, the stimulus–response event is necessarily an interaction of a human's sensory systems with the physical environment (i.e., the product or stimulus), and so psychophysics and sensory evaluation are essentially studying the same phenomena.

The reader should be careful not to confuse the stimulus with the sensation or response. This dichotomy is critical to understanding sensory function: an odor does not cause a smell sensation, but an odorant does. Thus, an odor is an experience and an odorant is a stimulus. Human observers or panelists do not measure sugar concentrations. They report their sweetness experiences. We often get into trouble when we confuse the response with the stimulus or vice versa. For example, people often speak of a sweetener being "200 times as sweet as sucrose" when this was never actually measured. It is probably impossible for anything to be 200 times as intense as something else in the sense of taste – the perceptual range is just not that large. What was actually measured was that, at an iso-sweet level (such as the threshold), it took a concentration 200 times higher to get the same impact from sucrose (or 1/200th for the intensive sweetener in question). But this is awkward to convey, so the industry as adopted the "*X* times sweeter than *Y*" convention.

The reader should also be careful to distinguish between the subjective experience, or sensation itself, and the response of the observer or panelist. One does not directly translate into the other. Often, the response is modified by the observer even though the same sensation may be generated. An example is how a stimulus is judged in different contexts. So there really are at least two processes at work here: the psychophysical event that translates energy into sensation (the subjective and private experience) and then the judgment or decision process, which translates that experience into a response. This second process was ignored by some psychophysical researchers, who assumed that the response was always an accurate translation of the experience. Figure 1.1 shows the two schema.

The rest of this chapter will look at various classical psychophysical methods, and how they are used to study absolute and difference thresholds. Some early psychophysical theory, and its modern variations, will also be discussed. A complete review of classical psychophysics can be found in *Psychophysics, The Fundamentals* (Gescheider, 1997).

## 1.2 Absolute Sensitivity

### 1.2.1 The Threshold Concept

In the early 1800s, at about the time when scientists were becoming better able to control the energy in a stimulus (light, sound, heat), a popular notion took hold that there was an energy level below which a human sensory system would not be able to detect anything, and above which a sensation would be noticed. This idea is often attributed to the philosopher Herbart, who wrote in 1824 that mental events had to be stronger than some critical amount in order to be consciously experienced (Gescheider, 1997). Thus, at any given moment, for a single observer, a stimulus with energy below the threshold is not detected, and a stimulus above the threshold is. This is an all-or-nothing concept of threshold, as shown in Figure 1.2.

As appealing as this concept seemed, it was almost impossible to demonstrate. Attempts to measure the threshold soon encountered a major problem: the sensitivity of the observer seemed to change from moment to moment. Although one might still like the idea that at any moment there was a fixed threshold, and that crossing it caused a sensation, attempts to bring the concept into the laboratory rendered the idea questionable, and of limited utility. Let us assume you are doing a study with changing sound pressure levels and that you ask your observer to respond "yes" when they hear something and "no" when they do not. We present all the stimulus tones of different levels many times in random order. When we plot the percentage of "yes" responses against the sound pressure levels, we do not see the all-or-nothing function of Figure 1.2, but rather a curve resembling an ogive or sigmoid shape, as shown in Figure 1.3a.

This was the first **psychometric function**, a term used to describe the probability of response as a function of the stimulus, such as its energy or sound pressure level. From this time forward, most people conceived of a threshold in practical terms, as a statistical entity rather than a fixed point. As a practical matter then, it became useful to call this empirical or experimental threshold the level at which detection occurs 50% of the time. This fits well with what we now know about the human nervous system: first, that it has a spontaneous

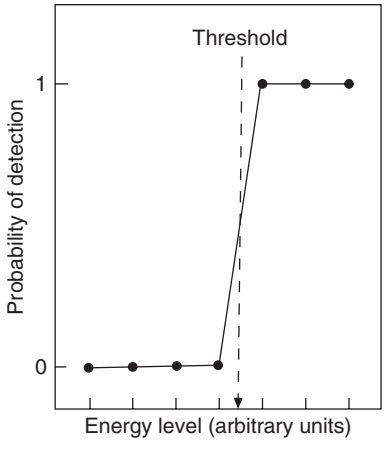


**Figure 1.2** The all-or-none concept of threshold. Detection never occurs below threshold. Above threshold, it always reaches consciousness.

background level of activity (your nerves are not quiet, even in a completely dark soundless room) and that this activity level appears to vary randomly.

As the normal distribution is well understood, has known properties, and describes a host of natural phenomena, it is perhaps not surprising that this curve is often used to describe the common psychometric function. As you might recall from statistics, the function has a location parameter (the mean $\mu$) and a spread parameter (the standard deviation $\sigma$ or variance $\sigma^2$). The exact form that generates our familiar bell shaped curve is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \tag{1.1}$$

And if we standardize the function to $\sigma = 1$ and $\mu = 0$ (we will use $\phi$ instead of $f$) we get

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \tag{1.2}$$

This is the equation, then, that gives us the relationship for *Z*-scores, or the conversion of any observation $x$ into its distance from the mean in standard deviation units as $Z = (x - \mu)/\sigma$. If we integrate this function over a given interval, we can find the cumulative proportion of the area under the curve from negative infinity to the upper bound of the interval. This gives us a useful relationship of proportions to *Z*-scores, as shown in Appendix 1.A. Taking our data from the hypothetical psychometric function, we can convert the proportions to *Z*-scores, as shown in Table 1.1. Plotting *Z*-scores instead of proportions will give us a linear relationship, amenable to least-squares fitting or other simple curve-fitting methods. Using our *Z*-score conversion, we see the linearization of the data from our curve in Figure 1.3b. The reader should feel comfortable with the conversion of proportions to *Z*-scores and vice versa, as this relationship will form the basis for several psychophysical calculations later in the book. Appendix 1.A illustrates these relationships.

Because the data are actually describing a probabilistic set of events that are bounded by zero and one, another attractive option is to use the logit function, which is similar to the

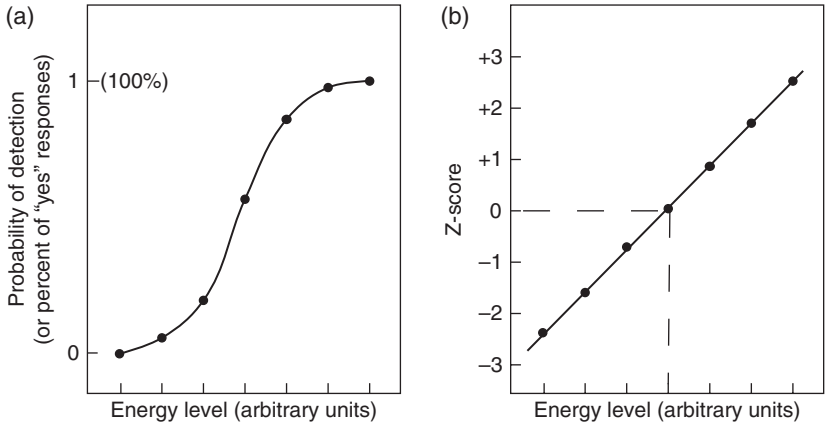


**Figure 1.3** A psychometric function. The probability of response is plotted against the stimulus energy level, and often forms an S-shaped curve similar to the cumulative normal distribution.

**Table 1.1** Conversion of threshold curve proportions to *Z*-scores

| Energy level | Proportion detecting | *Z*-score |
|---|---|---|
| 1 | 0.01 | –2.33 |
| 2 | 0.05 | –1.64 |
| 3 | 0.20 | –0.84 |
| 4 | 0.50 | 0.0 |
| 5 | 0.70 | +0.84 |
| 6 | 0.95 | +1.64 |
| 7 | 0.99 | +2.33 |

cumulative standard normal curve. It is often used to model probabilistic events in medicine and actuarial science, such as the rate of population growth and life expectancy. It takes the following form:

$$f(x) = \frac{e^x}{e^x + 1} = \frac{1}{1 + e^{-x}} \tag{1.3}$$

where $x$ can take on any value from negative to positive infinity and $f(x)$ varies from zero to one. If we convert our probability of detection $f(x)$ to an odds ratio $(p/1-p)$ and take the logarithm, the model becomes roughly linear and amenable to least-squares or other simple fitting methods:

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 x \tag{1.4}$$

where $b_0$ and $b_1$ are slope and intercept parameters. Additional terms (e.g., quadratic) may be added for greater accuracy. Examples of fitting logistic functions to threshold data can be found in Appendix 1.B, as well as in Walker et al. (2003) and Lawless (2010). The function is generally useful for modeling binomial events where there are two outcomes, such as correct versus incorrect responses in a choice task, or response versus nonresponse in a detection experiment.

Whichever bell-shaped curve we choose to use, and its cumulative S-form one chooses to adopt, it is clear that the mean and variance (location and spread parameters) of the distributions will affect the position of the threshold and the steepness of the S-curve. Figure 1.4 shows how the distributions of variability will affect the resulting psychometric function. Observers A and B have the same variability, but A has greater sensitivity and thus a lower threshold. Observers B and C have the same sensitivity, but observer C has less neural noise or other factors contributing to variance, and the smaller variability leads to a steeper psychometric function.

### 1.2.2 Threshold Theories

There were several theories that arose to explain this behavioral phenomenon of the apparent probabilistic nature of the threshold. One hypothesized that the actual momentary threshold of the observer was a fixed quantity, but that this varied according to a normal distribution. If so, over many trials, the data would look like the cumulative version of the normal distribution. Classical psychophysics calls this the phi–gamma

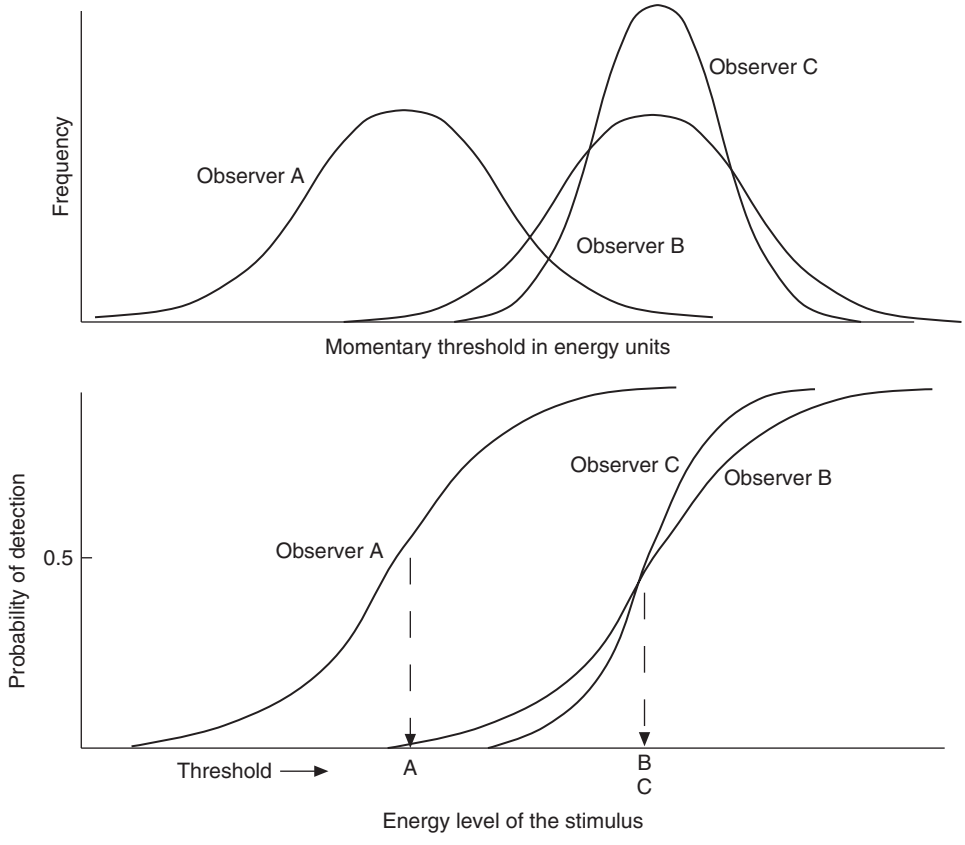


**Figure 1.4** Momentary and obtained thresholds from three hypothetical observers. Observer A has higher sensitivity than B or C because this person responds at a lower energy level. Observer C has lower inherent variability than Observer B, giving a steeper psychometric function.

hypothesis (Gescheider, 1997). Some workers held that if Fechner's log function held true (as discussed below) then the psychometric function obtained would take the form of the normal ogive when plotted against the log of stimulus intensity (thus the phi–log-gamma hypothesis). An example is shown in Figure 1.3 and one from actual data for odor thresholds in Figure 1.8. A second theory held that the observer's threshold was actually fixed, but the stimulus itself had some random variation that was normally distributed. If so, you could obtain the familiar psychometric function as a result of stimulus variation even from a fixed, static observer. In reality, probably both sources of variability are in play in most studies. Even the output of carefully controlled and well-engineered olfactometers can produce a variable stimulus to the nose, as Cain (1977) found when he actually measured the output of his olfactometer with a gas chromatograph.

### 1.2.3 Other Types of Thresholds

The absolute or detection threshold was only one type of threshold considered in psychophysics. The **recognition threshold** describes the level of energy needed for the observer to correctly name or identify the stimulus. For example, in the sense of taste, the detection

threshold is often lower than the recognition threshold. A common response is "I know I can taste something, but I can't name it." At a slightly higher concentration, say of sodium chloride, the observer correctly discerns a salty taste. The **difference threshold**, discussed at length below, is the level of energy change needed for the observer to perceive that the stimulus has become stronger or weaker. Because some sensory systems are known to reach a maximum or to saturate, the **terminal threshold** is used to describe the energy level at which subsequent increases in physical intensity fail to create any concomitant increase in sensation strength. This is rarely studied because it is difficult to present high-intensity stimuli without invoking a pain response or other changes in the stimulus quality. Recently, Prescott et al. (2005) have come up with a useful notion of the **rejection threshold**. This is the level at which a taste or odor would be found objectionable in a food or beverage product. They recognized that taints or off-flavors are not always objectionable at low levels. The rejection threshold is measured by conducting preference tests at increasing concentrations, comparing a spiked sample containing the off-flavor with a control without the offending substance.

## 1.3 Methods for Measuring Absolute Thresholds

### 1.3.1 The Method of Limits

The most widely used method for measuring thresholds in classical psychophysics was probably the **method of limits**. In this method, the stimulus energy is raised or lowered until the response of the observer changes. In an ascending series, the stimulus energy is increased until the observer responds that they detect the stimulus (i.e., get a perceived sensation). Descending trials are alternately conducted and these trials continue until the stimulus is no longer sensed. Because the individual's sensitivity is variable from moment to moment, several ascending and descending series would be presented, as shown in Figure 1.5. After the response appears to stabilize, the empirical threshold is obtained from an average of the change points in the last few runs.

Many experimenters realized that the method of limits, although appealing in its systematic presentations, was not very efficient. A lot of trials could be wasted when the observer sensed the higher levels of the stimulus and perception was fairly obvious. Why not use the previous reversal point as a starting level for the next series? This notion led to various adaptive methods for measuring threshold. They are adaptive in the sense that the next series is adapted to take into account the information we get from the current series of stimuli. An example of this is the *staircase procedure*, so named because when the stimulus sequence is connected by horizontal lines the record resembles a rising and falling series of steps (Cornsweet, 1962; see also Linschoten et al. (1996) for a modern variant). In its simplest form, the stimulus is raised on the next trial if no sensation is detected (i.e., the observer responds negatively) and the stimulus is lowered on the next trial if the observer responds affirmatively, as shown in Figure 1.6. Many variations on this procedure were developed, such as a random double staircase in which trials would randomly be drawn from two intertwined staircases, one starting from a high level of stimulus intensity and the other from a low level. Another popular option was to descend only after two positive responses, but to ascend after one negative, the so-called up–down transformed-response rule (UDTR). This method titrated around a 71% detection level. These methods are discussed at greater length in Lawless and Heymann (2010: chapter 6), and the reader is referred there for further methodological details.

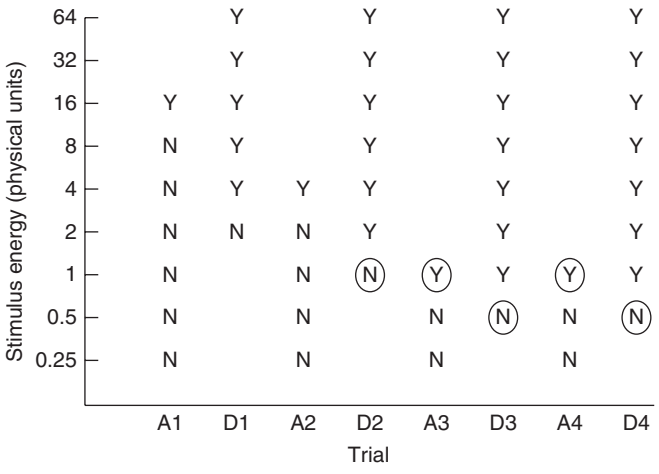


**Figure 1.5** The method of limits. Stimuli were presented in alternating ascending (A1, A2, etc.) and descending (D1, D2, etc.) series and the observer responds either "no" if they do not sense anything (N) or "yes" on each trial. After responses appear to stabilize, the last few points of change (circled) are averaged to obtain the threshold. In this case, we average the change points 1, 1, 0.5, 1, and 0.5, giving a mean of 0.8 as the individual's threshold estimate.

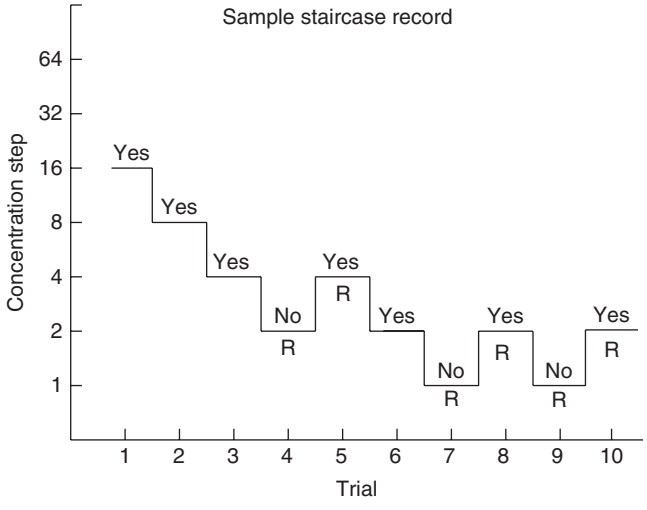


**Figure 1.6** An example (hypothetical) of a record from a staircase procedure.

### 1.3.2 Forced-Choice Method of Limits

Several problems remained with the classical method of limits. One was that the observer could become adapted or fatigued to the high levels of the stimulus in the descending series. For example, with a bitter substance, the taste is difficult to remove, and thus lower level trials became increasingly difficult to discern due to a buildup of residual bitter taste in the mouth. So, in the chemical senses, the method of limits is often performed with only ascending trials. The second problem is that the actual response is both a function of the sensitivity of the observer (what you are really trying to measure) and their criterion for how much of

a sensation is needed for a positive response. Some observers could be quite conservative and only respond when they are absolutely sure they heard or saw or tasted something, while another person might respond if they had even the slightest inkling that something was sensed. Thus, the individual's proclivity to respond or not respond with various levels of evidence is a confounding influence in the threshold we obtain. This issue of separating response bias or one's individual criterion from the actual sensitivity is addressed by signal detection theory, as explained in detail in Chapter 3.

A simple solution to this problem is to force the observer to prove to us that they can detect the stimulus by correctly choosing it from amongst a group of other stimuli that have only the background noise. Let us call the stimulus with the higher energy present the target, and the stimulus at the background level the blank. An example of a blank in the sense of taste would be a solution of distilled water that contained none of the tastant whose threshold you are trying to measure. Typically, this is done with two, three, or four total stimuli at each level, although a variety of numerical combinations of targets and blanks has been used in different sensory studies (e.g., Lawless et al., 1995). A good example of such a procedure is the **ascending forced-choice method of limits** described by ASTM procedure E-679 (ASTM, 2008). In this case, each observer in a group of from 10 to 25 individuals is given an ascending series with one target and two blank stimuli at each level, as is asked to choose the target item. The individual threshold estimate becomes the first correct response in the series, given that all subsequent responses are also correct (a correct choice followed by an incorrect choice is discounted as probably a lucky guess). Then the final reversal points are tabulated across the group of individuals to obtain a group estimate threshold for a particular taste or odor substance (see Stocking et al. (2001) for an example). Because the stimulus series is often in a geometric progression, such as concentration steps in factors of two or three, the geometric rather than the arithmetic mean is often taken. The geometric mean is the *N*th root of the product of *N* items. A convenient calculation for the geometric mean is to convert the data via logarithmic transformation, take the mean of the log data, and then the antilog of that mean, as shown in eqns 1.5–1.7.

$$\text{Mean}_{\text{geometric}} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \cdots \cdot x_n} \tag{1.5}$$

where $x_1$, $x_2$, and so on are the individual best-estimate thresholds from a panel of $n$ individuals.

We can also take the arithmetic mean of the logs (let us call it "ML") and then exponentiate to get the antilog:

$$\text{ML} = \frac{\sum_{i=1}^{n} \log x_i}{n} \tag{1.6}$$

Assuming our logarithms were to the base 10:

$$\text{Mean}_{\text{geometric}} = 10^{\text{ML}} \tag{1.7}$$

As a measure of central tendency, the geometric mean has the property (like the median) that it is less influenced by high outliers in the data than the arithmetic mean is. Also, because the data are in a geometric progression due to the doubling or tripling of concentration steps, they are in equal steps when converted to logarithms. This makes the data handling easy if you have to calculate geometric means by hand.

The *N*-alternative ascending forced choice method is a useful procedure. However, the data handling outlined above (i.e., by getting individual threshold estimates and then averaging them) still does not deal with the fact that the participants can guess correctly (and do so one-third of the time for the ASTM E-679 procedure). A solution to this problem is to use the common correction for guessing, sometimes called **Abbott's formula**, as shown in eqns 1.8 and 1.9:

$$P_D = \frac{P_{Obs} - P_{chance}}{1 - P_{chance}} \tag{1.8}$$

where $P_D$ is the true proportion detecting, $P_{obs}$ is the proportion choosing the target sample correctly, and $P_{chance}$ is the chance level or 1/3 for the ASTM method. This can also be recast as

$$P_{obs} = P_{chance} + P_D\left(1 - P_{chance}\right) = P_D + P_{chance}\left(1 - P_D\right) \tag{1.9}$$

One can think of the $P_{obs}$ in these equations as *the level you need to get to* in order to obtain your threshold after the correction; in other words, the adjusted percentage correct. Note that this assumes there are only two classes of individuals at any given level, a group that detects the stimulus and those that do not, some of whom guess correctly. This two-state model is sometimes attributed to Morrison (1978).

Now all we need to do to obtain the group threshold is to count the proportion correct at each concentration level and then interpolate, graphically or from a fitted equation. The interpolation point for any forced-choice method is that which satisfies Abbott's formula at 50% actual discriminators (see Antinone et al. (1994) for an example). We do not need to worry about estimating individual thresholds or whether a person made a lucky guess or not at each trial. Every person's data contributes equally. To obtain the historical criterion of 50% detection for threshold, we need to interpolate at 2/3 or 66.6% correct in the ASTM choose-one-out-of-three type of procedure. This interpolation is shown in Figure 1.7. A second advantage of this method of data analysis is that we can also estimate other levels of detection after correction for guessing. For example, we might want to know the level at which 25% of the population can detect a substance or even 10% (requiring 50% observed correct and 40% for the ASTM method, respectively). For a food scientist, trying to track down the source of an off-flavor, finding that your food has an offending chemical present at the consumer threshold at 50% is not too useful. One might want to limit the amount to insure that far fewer people can detect it, thus leading to fewer complaints and less loss of the business franchise due to consumers' dissatisfaction.

Figure 1.8 shows the method applied to the data of Stocking et al. (2001) on odor thresholds for methyl tertiary butyl ether (MTBE, a water contaminant). Using a simple linear fit of the equation shown, we get a threshold value of about 14 ppb (parts per billion). Fitting a logistic equation (plotting $\ln(p/1-p)$ versus log concentration), the interpolated value is about 12 ppb. These values agree reasonably well with the level estimated by the ASTM procedure of taking individual best estimates and then finding the geometric mean, yielding a value of 14.3 ppb in the Stocking et al. data. A method for fitting a logistic equation such as eqn 1.4 to the data is shown in Appendix 1.B. Another alternative for forced-choice data is given in the paper by Harwood et al. (2012) using a four-parameter logistic equation as follows:

$$P_{\text{observed}} = \text{Min} + \left[\frac{\text{Max} - \text{Min}}{1 + 10^{k(\log(T) - X)}}\right] \tag{1.10}$$

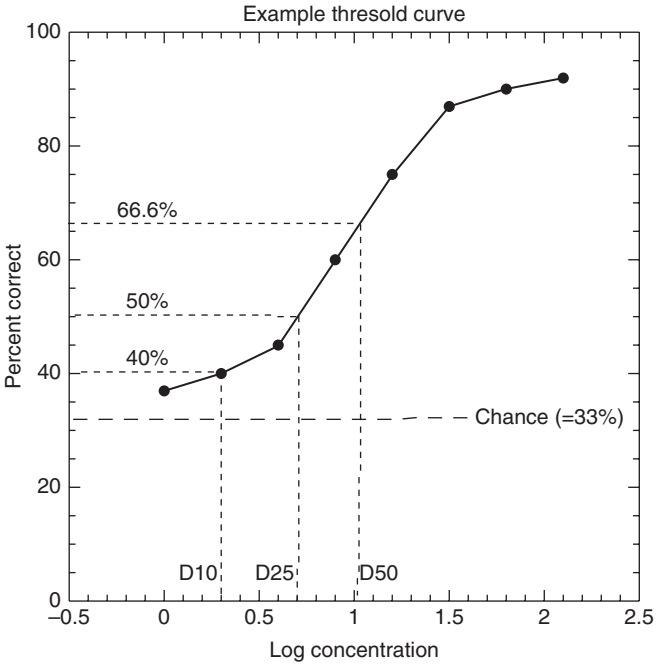


**Figure 1.7** Using the 3-AFC method of ASTM E-679, we can interpolate on a plot of proportion correct versus concentration (or log concentration) to find the chance-corrected threshold level (66.6% correct for 50% detection) and also other levels; for example, for 25% detection (in this case 50% correct) and 10% detection (40% correct).

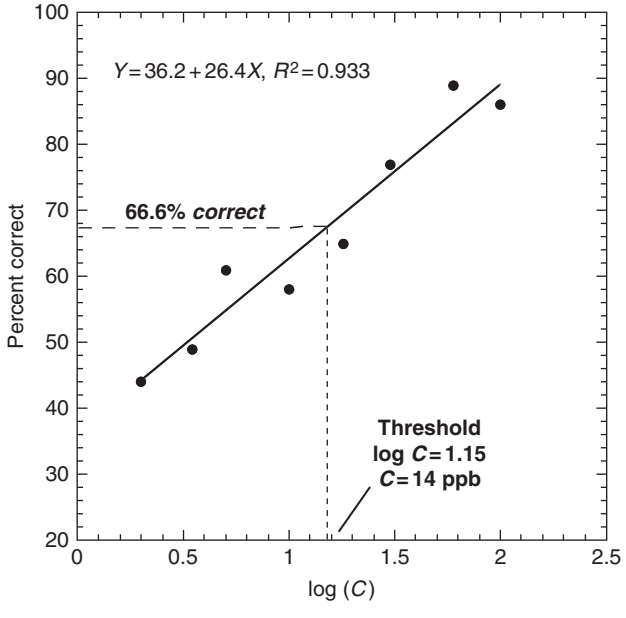


**Figure 1.8** An example of interpolation on the percentage correct from a three-alternative forced-choice test to obtain the chance-corrected 50% discrimination level, in this case using the data of Stocking et al. (2001). Using the equation shown, we get a threshold of about 14 ppb; and using a logistic regression, we get about 12 ppb.

where Min and Max are the minimum and maximum levels of response, $k$ is a slope parameter, $\log(T)$ is the chance corrected threshold level to be found, and $X$ is the log concentration of the stimulus. The values for Max and Min are generally known from the data, reducing the solution to a search for values of $\log(T)$ and $k$. Another fitting method for these kinds of data, using maximum likelihood methods, is found in Peng et al. (2012: appendix A).

Yet one problem still remains. We have changed the definition of threshold when we get to a group-averaged or population measure, when we speak of 50% *of the group* detecting. Remember that when we began our discussion of classical thresholds we talked about a single observer detecting 50% *of the time*. While reanalyzing the data of Stocking et al. (2001), a statistician from the US Environmental Protection Agency, Andrew E. Schulman, recognized this embedded issue that was being overlooked in the world of practical threshold estimation (USEPA, 2001). He stated (USEPA, 2001: 49):

> Odor detection thresholds should be defined as the concentrations at which a certain percent of people can detect the contaminant a certain percent of the time. Both the time and subject fractions must be specified in order for a threshold to be interpretable.

After extensive statistical modeling of the Stocking et al. (2001) data, he concluded that the ASTM method of analysis appeared to find the point at which 50% of the group would detect 50% of the time. Unfortunately, if Schulman's analysis is correct, the overall probability of any detection event would thus be 0.25 rather than our classical definition of 0.50. Subsequent researchers will probably need to sort out this issue in future methodological studies.

Many modifications of the ascending forced-choice methods can be found in the literature. One important decision is whether to invoke a **stopping rule**. A stopping rule allows the researcher to terminate the ascending series after a sequence of correct answers, typically three in a row. This prevents the panelist or subject from experiencing very high levels of the stimulus that could be unpleasant, painful, or cause fatigue or adaptation as the experimental series progresses. The exact choice for the stopping rule, as well as the definition of threshold, will influence the values obtained (Peng et al., 2012).

## 1.4 Differential Sensitivity

### 1.4.1 The Difference Threshold

The second most important psychophysical phenomenon in classical psychophysics was the different threshold. The difference threshold is the minimum amount of stimulus change (in energy or physical units) that is necessary for the observer to note that the stimulus has become stronger or weaker. Other terms for this concept include the difference limen (DL) and the just-noticeable difference (JND). In some ways the DL was of greater practical utility than the absolute threshold, because it dealt with sensations across the full stimulus range that the observer could respond to. The entire dynamic range of stimulus and response could now be measured and quantified, and not just the very weak sensations around threshold. In comparing stimuli to measure a difference threshold, we shall call the baseline or starting stimulus the "standard" stimulus and the second stimulus, whose level will be varied, the "comparison" stimulus.

**Table 1.2** Examples of Weber fractions for different sensory modalities*

| Sensory system or modality | Weber fraction |
|---|---|
| Electric shock | 0.013 |
| Saturation, red | 0.019 |
| Heaviness | 0.020 |
| Finger span | 0.022 |
| Length | 0.029 |
| Vibration, 60 Hz | 0.036 |
| Loudness | 0.048 |
| Brightness | 0.079 |
| Taste, NaCl | 0.083 |

*From Teghtsoonian (1971).

In the 1830s, the German physiologist E.H. Weber studied the size of the difference threshold, mostly working with lifted weights. He found that, as the stimulus became heavier, it took a larger increase in weight to be just-noticeably heavier. In other words, heavier weights became harder to discriminate or to tell apart (Stevens, 1971; Gescheider, 1997). But the truly valuable insight came when he looked at the size of the difference threshold relative to the weight of the standard stimulus. The ratio of the stimulus change to the level of the standard stimulus was virtually constant. If we let $\Delta E$ be the size of the change necessary to produce a JND and $E$ be the energy (or weight or concentration) of the standard, then the following relationship held:

$$\frac{\Delta E}{E} = c \quad \text{and thus} \quad \Delta E = cE \tag{1.11}$$

In other words, the weights had to change by a constant proportion or percentage $c$ in order to cross the perceivable boundary. The DL or JND was found to be a constant fraction or proportion of the starting level. This was the first quantified psychophysical relationship and has become known as **Weber's law** (Stevens, 1971; Gescheider, 1997). The Weber fraction $c$ could be used as a measure of the resolving power of any sensory system or modality. Researchers could apply the techniques of psychophysics to compare the functioning of different sensory modalities. It became apparent, for example, that the visual and auditory senses were able to discriminate much smaller percentage changes in a stimulus energy level, than touch, taste, or olfaction senses could. Some examples are shown in Table 1.2. Measuring the DL and determining the Weber fraction for different senses and different conditions would keep graduate students busy for many decades to come. Fechner was once touted as having made no less than 24,576 judgments in testing Weber's law for lifted weights (James, 1892).

But how general was Weber's "law"? A common observation was that at very low stimulus levels, those approaching the absolute threshold, that the Weber fraction increased (Stevens, 1971). That is, it took a larger percentage change at very low levels to be detectably different than the change that was needed over the middle of the functional stimulus range. Sometimes this departure (increase) was also noted at high levels of the stimulus as well. So when $\Delta E/E$ was plotted as a function of $E$, it did not always form a flat horizontal

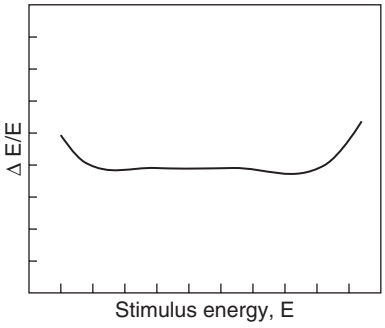


**Figure 1.9** The theoretical relationship of Weber's law and the common finding of departures at low and high levels of energy.

line, but might curve up at the ends as in Figure 1.9. In order to reconcile this discrepancy, sometimes an additive constant was included in Weber's law:

$$\frac{\Delta E}{E+k} = c \quad \text{and thus} \quad \Delta E = c\left(E+k\right) \tag{1.12}$$

This constant $k$ adjusted for the curvature of the Weber fraction plot at the lower end of the range. The exact meaning of this convenient constant is not clear, but it could represent some background noise in the sensory system that becomes more important at low levels around threshold measurement, and less of a factor as the stimulus level becomes stronger (Gescheider, 1997). Note that the additive constant does not help with the breakdown of Weber's law at very high levels. However, working at high stimulus levels is not easy, and other factors come into play. For example, a very strong stimulus may become painful, and thus the stimulus is now multidimensional in terms of the sensations it is causing.

Note that the absolute threshold can also be conceived of as a special case of the difference threshold. When we experimentally determine an absolute threshold, we are usually making a comparison with some blank, background, or "pure noise" stimulus, such as distilled water in the case of taste studies. Thus, the detection or absolute threshold is empirically the difference threshold against this neutral background.

### 1.4.2 Methods for Measuring Difference Thresholds

The common method for measuring difference thresholds was to present a long sequence of paired comparisons in which one member of the pair was held constant (the standard stimulus) and the second (the comparison stimulus) was varied, usually both above and below the standard. This was the "bread and butter" of classical psychophysics when studying Weber fractions, and was known as the **method of constant stimuli**. The term is perhaps a bit unfortunate, because only one item is really constant. The task for the observer is to say whether the comparison stimulus is heavier (in the case of lifted weights) or lighter than the standard. Note that you can also conduct this experiment as a blind comparison, and just ask the observer to choose which item is heavier. A choice is forced, although some variations of the method allow for an "equal" judgment as well. A random order of comparison stimuli in the pairings is presented; that is, there is no fixed order, increasing or decreasing, of the comparison items.

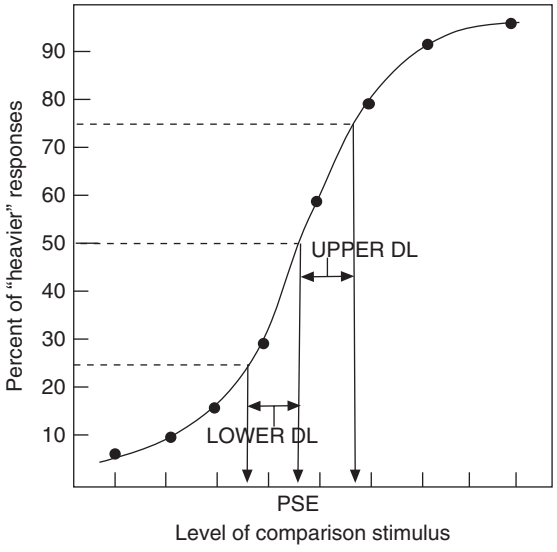


**Figure 1.10** The method of constant stimuli. The standard stimulus is a fixed value and the comparison stimulus is varied around it. In an experiment on lifted weights, the observer may be forced to respond either "heavier" or "lighter" when comparing the comparison item with the standard. A plot of the proportion "heavier" will give an upper and a lower difference threshold, found by interpolating at the 75% and 25% points, respectively. A sigmoid or ogive curve is a common observation.

Figure 1.10 shows the critical plot of the data. We consider only one of the responses, in this case "heavier." When the comparison is lower in weight than the standard, the percentage of "heavier" responses will usually be below 50% and drop off as the comparison weight decreases. Conversely, the percentage rises above 50% as the comparison is physically greater in weight. Upper and lower difference thresholds (DLs) can be determined by the physical change necessary to produce a 75% and 25% response, and taking their difference from the 50% point. Note that the DL is stated in physical units.

Often, the 50% point does not exactly correspond to the standard stimulus. This was known as the **point of subjective equality** (PSE), and the difference from the physical value of the standard was know as the **constant error** (CE).

The method of limits could also be adapted to the measurement of difference threshold. The sequence of pairs would now be increasing or decreasing, rather than random. In modern times this would be considered less than optimal, because it would very likely induce some expectations on the part of the observer that things are going to change in a certain direction. It was also common to allow the observer to respond "equal." This introduces a further problem, in that the observer has to decide what level of change is necessary to leave the comfortable "equal" option. A conservative observer might want to be sure that the comparison was different, while a more lax observer might be inclined to respond with very little perceived difference. So there is a criterion problem here again, which will be dealt with in Chapter 3.

Figure 1.11 shows how the data might look from a hypothetical experiment on lifted weights. The comparison stimulus is changed to move it toward the standard in ascending and descending runs. Note that the run continues past the level of the standard. At some

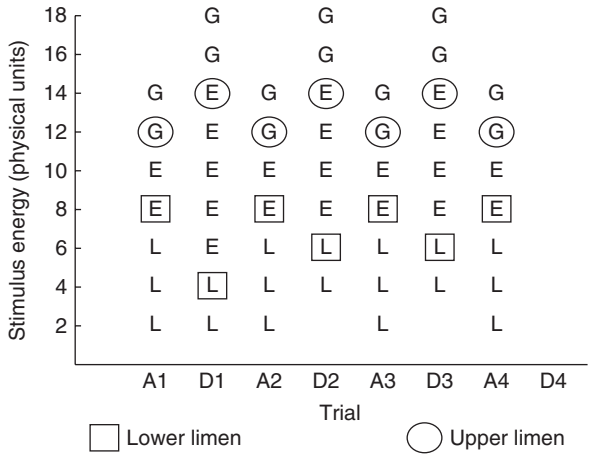


**Figure 1.11** The method of limits with an "equal" response used to obtain the difference threshold. The IU is found from the average of the lower limens subtracted from the average of the upper limens, the points at which the response changes. The DL is then one-half the IU.

point, while increasing from a low level, the observer's response changes from "lighter" to "equal" and, as the series progresses, from "equal" to "heavier." Thus, there are two change points in each series, and these define the **interval of uncertainty** (IU). The upper change point was deemed the upper limen or $L_u$ and the lower change point was the lower limen or $L_l$. After several series, the mean lower limen would be subtracted from the mean upper limen to get the IU and then the difference threshold would be one-half of the IU. The PSE could also be defined as the midpoint of the two means.

## 1.5 A Look Ahead: Fechner's Contribution

G.T. Fechner is considered the father of psychophysics, largely due to the publication of his book on the topic in 1860. Fechner had studied medicine, mathematics, and physics and later in his life turned to philosophy. He was troubled by the mind–body dichotomy of René Descartes, and felt that mind and matter were equal in the sense they were two manifestations of the same reality. On 22 October in 1848, or so the story goes, Fechner had the insight that Weber's law would connect the two. He was also interested in finding a way to represent subjective experience; that is, the loudness of a sound or the brightness of a light. Until this time, we only knew how much to change the light or the sound to create a sensation or a sensation difference.

The problem of providing a numerical representation of subjective magnitude could be solved, according to Fechner, if one used the JND as the unit of subjective increase. Assuming that all JNDs were subjectively equal, one merely had to add them up (starting with absolute threshold) to provide a measure of how strong a sensation appeared to the observer. A plot of the JNDs would provide the ruler needed to define a given subjective magnitude for some stimulus continuum, as shown in Figure 1.12. Note that this would only hold over the range in which Weber's law was accurate. Fechner was also familiar with the branch of mathematics we have come to call calculus, and understood that, as one accumulated a larger and larger number of smaller and smaller intervals, one could integrate

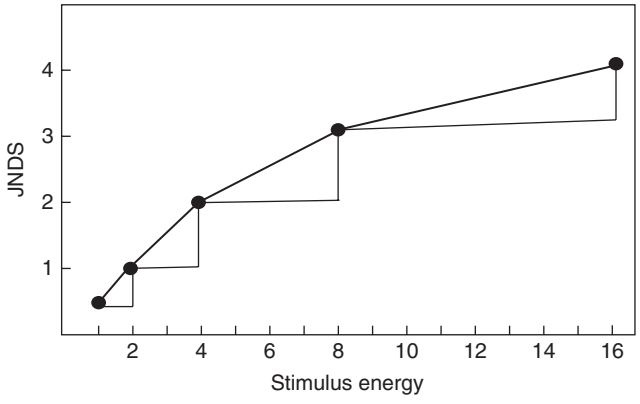


**Figure 1.12** A plot of JNDs is accumulated to obtain the psychophysical function relating subjective intensity against physical stimulus intensity, as according to Fechner.

to get this summed relationship. Thus, since the integral of d$x/x$ is the natural logarithm of $x$ (plus a constant of integration), Fechner's law was given as a logarithmic relationship between subjective intensity $S$ and stimulus energy $E$:

$$S = k \log E \tag{1.13}$$

This logarithmic relationship was considered a psychophysical "law" and is sometimes combined with Weber's law to be called the Weber–Fechner relationship. As a rule of thumb, it is not a bad idea to keep in mind. It tells us that, in order to make equal subjective changes in sensation strength, we need to vary the stimulus strength in a geometric progression, and not equal arithmetic steps. If I want to increase the sweetness of a beverage by equal steps, I had better not use 8%, 10%, and 12% sucrose. The second jump will probably be smaller than the first. A proportional change across the steps, for example, by a multiplicative factor of 1.2, would be more likely to produce equal steps of sweetness increases. As our quote from Shakespeare at the outset of the chapter shows, the candle is overshadowed in comparison with the light of the moon. The rule is, as the stimulus level increases, it generally takes larger steps to notice a difference. Against the background of a dark night, the candle is observed, but not against the light of the shining moon.

## Appendix 1.A Relationship of Proportions, Areas Under the Normal Distribution, and *Z*-Scores

Because the exact shape of the normal distribution is known, the area under the curve from $-\infty$ to $+\infty$ to any value (usually expressed in *Z*-scores) can be estimated. Thus, there is a simple relationship between area or probability $p$ and any value's distance from the mean, expressed in standard deviation units (*Z*-scores). The relationship between $p$ and $Z$ is shown in Table 1.A.1 and in Figure 1.A.1.

**Table 1.A.1** Proportions and *Z*-scores*

| Proportion | *Z*-score | Proportion | *Z*-score | Proportion | *Z*-score | Proportion | *Z*-score |
|---|---|---|---|---|---|---|---|
| 0.01 | −2.33 | 0.26 | −0.64 | 0.51 | 0.03 | 0.76 | 0.71 |
| 0.02 | −2.05 | 0.27 | −0.61 | 0.52 | 0.05 | 0.77 | 0.74 |
| 0.03 | −1.88 | 0.28 | −0.58 | 0.53 | 0.08 | 0.78 | 0.77 |
| 0.04 | −1.75 | 0.29 | −0.55 | 0.54 | 0.10 | 0.79 | 0.81 |
| 0.05 | −1.64 | 0.30 | −0.52 | 0.55 | 0.13 | 0.80 | 0.84 |
| 0.06 | −1.55 | 0.31 | −0.50 | 0.56 | 0.15 | 0.81 | 0.88 |
| 0.07 | −1.48 | 0.32 | −0.47 | 0.57 | 0.18 | 0.82 | 0.92 |
| 0.08 | −1.41 | 0.33 | −0.44 | 0.58 | 0.20 | 0.83 | 0.95 |
| 0.09 | −1.34 | 0.34 | −0.41 | 0.59 | 0.23 | 0.84 | 0.99 |
| 0.10 | −1.28 | 0.35 | −0.39 | 0.60 | 0.25 | 0.85 | 1.04 |
| 0.11 | −1.23 | 0.36 | −0.36 | 0.61 | 0.28 | 0.86 | 1.08 |
| 0.12 | −1.18 | 0.37 | −0.33 | 0.62 | 0.31 | 0.87 | 1.13 |
| 0.13 | −1.13 | 0.38 | −0.31 | 0.63 | 0.33 | 0.88 | 1.18 |
| 0.14 | −1.08 | 0.39 | −0.28 | 0.64 | 0.36 | 0.89 | 1.23 |
| 0.15 | −1.04 | 0.40 | −0.25 | 0.65 | 0.39 | 0.90 | 1.28 |
| 0.16 | −0.99 | 0.41 | −0.23 | 0.66 | 0.41 | 0.91 | 1.34 |
| 0.17 | −0.95 | 0.42 | −0.20 | 0.67 | 0.44 | 0.92 | 1.41 |
| 0.18 | −0.92 | 0.43 | −0.18 | 0.68 | 0.47 | 0.03 | 1.48 |
| 0.19 | −0.88 | 0.44 | −0.15 | 0.69 | 0.50 | 0.94 | 1.55 |
| 0.20 | −0.84 | 0.45 | −0.13 | 0.70 | 0.52 | 0.95 | 1.64 |
| 0.21 | −0.81 | 0.46 | −0.10 | 0.71 | 0.55 | 0.96 | 1.75 |
| 0.22 | −0.77 | 0.47 | −0.08 | 0.72 | 0.58 | 0.97 | 1.88 |
| 0.23 | −0.74 | 0.48 | −0.05 | 0.73 | 0.61 | 0.98 | 2.05 |
| 0.24 | −0.71 | 0.49 | −0.03 | 0.74 | 0.64 | 0.99 | 2.33 |
| 0.25 | −0.67 | 0.50 | 0.00 | 0.75 | 0.67 | 0.995 | 2.58 |

*Calculated in Excel®.

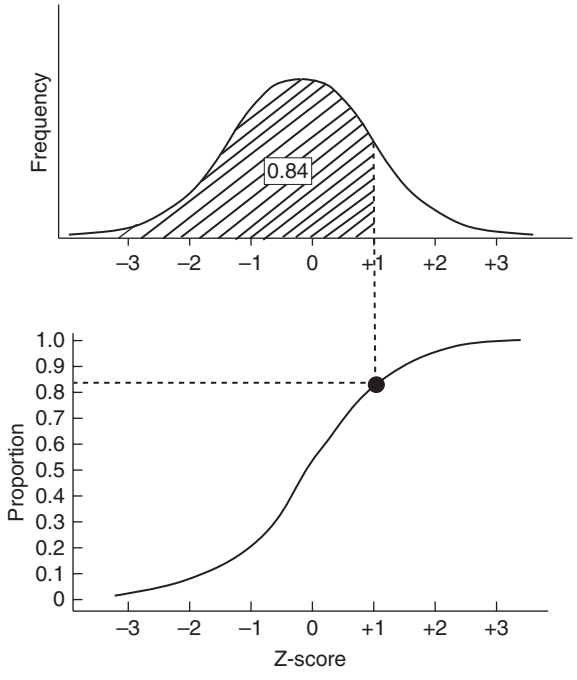


**Figure 1A.1** Each point on the ogive curve corresponds to an area under the normal curve, or proportion of the total area that is below and to the left of that point.

## Appendix 1.B Worked Example: Fitting a Logistic Function to Threshold Data

In this example, we will take the data from Stocking et al. (2001), fit a logistic function using ordinary least squares, and then interpolate to find the chance-corrected 50% detection level. Recall that Stocking et al. used the ASTM E-679 procedure, which is a hybrid of a 3-AFC and triangle test. It resembles the 3-AFC in that only one of the three samples contains the odorant, in this case MTBE. It resembles a triangle test in that participants are asked to choose the odd sample, rather than the strongest smelling. The actual strategy adopted by testers is usually unknown. Stocking et al. used 57 testers and eight concentration steps, differing roughly by a factor of 2. The data that we will use consists of the percentage of correct choices at each concentration level. The actual ASTM procedure requires finding individual thresholds based on some heuristic rules, and then averaging, but this discounts some correct judgments, so we will use the entire data set and percentages correct. The data set can be found in Stocking et al. (2001) and also in the appendix to Chapter 6 in Lawless and Heymann (2010).

Recall that for a three-alternative procedure, we need to apply the correction for guessing, also known as Abbott's formula, in order to find the true proportion detecting using the following expression:

$$P_{\mathrm{D}} = \frac{P_{\mathrm{obs}} - P_{\mathrm{chance}}}{1 - P_{\mathrm{chance}}} \tag{1.B.1}$$

where $P_{\mathrm{D}}$ is our true percentage detecting and $P_{\mathrm{obs}}$ is the proportion observed in the data that would yield the desired $P_{\mathrm{D}}$. With a chance probability of 1/3, we require 2/3 correct to get a true detection rate of 0.5, 50% correct to get a detection rate of 0.25, and 40% correct to get a true proportion of 0.10. After we fit our equation to the data, we will interpolate at these three levels. In this example, we will fit the function to the raw proportion correct, fit the function, and then interpolate at the corrected levels. We could have corrected the proportion before fitting as well, but this does not seem to affect the results very much (USEPA, 2001).

The logistic function takes the form of

$$f(x) = \frac{\mathrm{e}^{x}}{\mathrm{e}^{x} + 1} = \frac{1}{1 + \mathrm{e}^{-x}} \tag{1.B.2}$$

We can "linearize" this relationship by converting our proportion correct to an odds ratio, $p/(1-p)$:

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 x \tag{1.B.3}$$

Table 1.B.1 shows the raw proportion correct, the odds ratio (our *Y* variable), the concentration steps in parts per billion, and the log of the concentrations (our *X* variable). The logarithm of concentration is used because there is a roughly geometric progression in the concentration steps.

The method of least squares, or ordinary least squares, fits a straight line to the data by minimizing the squared residual deviations in the *Y*-direction of the actual data points to the

**Table 1.B.1** Proportions correct versus log concentration from the Stocking et al. (2001) data

| Conc. | log(*c*) "*X*" | Prop. *p* | *p*/(1 – *p*) | ln[*p*/(1 – *p*)] "*Y*" | $X^2$ | $Y^2$ | *XY* |
|---|---|---|---|---|---|---|---|
| 2 | 0.301 | 0.44 | 0.786 | –0.241 | 0.091 | 0.058 | –0.072 |
| 3.5 | 0.544 | 0.49 | 0.961 | –0.039 | 0.296 | 0.002 | –0.022 |
| 6 | 0.778 | 0.61 | 1.564 | 0.447 | 0.606 | 0.199 | 0.348 |
| 10 | 1.000 | 0.58 | 1.381 | 0.323 | 1.000 | 0.104 | 0.322 |
| 18 | 1.255 | 0.65 | 1.857 | 0.618 | 1.576 | 0.383 | 0.777 |
| 30 | 1.477 | 0.77 | 3.348 | 1.208 | 2.182 | 1.460 | 1.784 |
| 60 | 1.778 | 0.89 | 8.091 | 2.091 | 3.162 | 4.371 | 3.717 |
| 100 | 2.000 | 0.86 | 6.143 | 1.815 | 4.000 | 3.295 | 3.631 |
| Sum | 9.133 | | 24.131 | 6.223 | 12.91 | 9.874 | 10.49 |
| Mean | 1.142 | | | 0.778 | | | |

intercepted value on the fit line. Although there are many ways to fit a function to the data, this is probably the most commonly used method.

The sum of the squared residual deviations $\sum R^2$ will be minimized when the partial derivatives of $R$ with respect to the slope $b$ and intercept $a$ are set to zero as follows:

$$y = a + bx \tag{1.B.4}$$

$$\sum R^2(a,b) = \sum_{i=1}^{N} [y_i - (a + bx_i)]^2 \tag{1.B.5}$$

And setting the partial derivatives to zero:

$$\frac{\partial R^2}{\partial b} = -2\sum_{i=1}^{N} [y_i - (a + bx_i)]x_i = 0 \tag{1.B.6}$$

and

$$\frac{\partial R^2}{\partial a} = -2\sum_{i=1}^{N} [y_i - (a + bx_i)] = 0 \tag{1.B.7}$$

Solving the two equations in two unknowns yields the following expressions for $a$ and $b$ (dropping the summation index as we are always summing from 1 to $N$ data points:

$$b = \frac{N\sum xy - \left(\sum x\right)\left(\sum y\right)}{N\sum\left(x^2\right) - \left(\sum x\right)^2} \tag{1.B.8}$$

$$a = \frac{\left(\sum x^2\right)\sum y - \sum x \sum xy}{N\sum x - \left(\sum x\right)^2} \tag{1.B.9}$$

These become a little simpler if we work in the means of $x$ and $y$ (using the bar symbol for the mean) as follows:

$$b = \frac{\sum xy - N\overline{xy}}{\sum\left(x^2\right) - N\left(\overline{x}\right)^2} \tag{1.B.10}$$

**Table 1.B.2** Summary of the procedure

| Percentage detecting | Percentage correct required after correction | Converting to $\ln[p/(1-p)]$ | Solving for $\log(c)$ via $x=(y-a)/b$ | $10^x$ (ppb) |
|---|---|---|---|---|
| 50 | 66.7 | 0.693 | 1.079 | 12.0 |
| 25 | 50.0 | 0.0 | 0.571 | 3.7 |
| 10 | 40.0 | –0.405 | 0.274 | 1.9 |

and

$$a = \overline{y} - b\overline{x} \tag{1.B.11}$$

Returning to the Stocking et al. data, we can now make our calculations using eqns 1.B.10 and 1.B.11:

$$b = \frac{10.49 - 8(1.142)(0.778)}{12.91 - 8(1.142)^2} = \frac{3.382}{2.478} = 1.364$$

and

$$a = 0.778 - 1.364\left(1.142\right) = -0.779$$

Now we can solve for $\log(c)$ for our three points of interest. Solving the equation for $x$ gives us $x=(y–a)/b$, so $\log(c)=(y–0.778)/1.364$. Our proportions of detectors again are 0.667, 0.5, and 0.4. Converting to the natural log of the odds ratio, $\ln[p/(1-p)]$, gives us $y$ values of 0.693, 0.0, and –0.405 for $y$-values. Our interpolations then are as follows:

- $[0.693-(-0.779)]/1.364=1.079$, and so $10^{1.079}=12.0$ ppb for our threshold estimate;
- for our 25% detection rate, we get $[0-(-0.779)]/1.364=0.571$ and $10^{0.571}=3.7$ ppb;
- and for our 10% detection level we get $[-0.405-(0.779)]/1.364=0.274$ and $10^{0.274}=1.9$ ppb.

The procedure is summarized in Table 1.B.2.

These estimates are actually quite close to those you can obtain via the ASTM method, which gives a threshold in the range of 14–15 ppb, reasonably close to our estimate of 12 ppb via curve fitting and interpolation. Simple inspection of the raw data also provides some validation. For example, there were 10 people who chose the correct sample all the way through the data set and who might have been detecting the MTBE sample at the lowest level of 2 ppb. If we take that proportion of 10/57 we get about 17.5%. Correcting this for guessing gives us about 10% as a likely estimate for true detectors at the 2 ppb level, which is quite close to our interpolated estimate of 1.9 ppb!

## References

Antinone, M.A., Lawless, H.T., Ledford, R.A., and Johnston, M. 1994. The importance of diacetyl as a flavor component in full fat cottage cheese. Journal of Food Science, 59, 38–42.

ASTM. 2008. Standard practice for determining odor and taste thresholds by a forced-choice ascending concentration series method of limits, E-679-04, Annual Book of Standards, Vol. 15.08. ASTM International, Conshocken, PA, pp. 36–42.

Cain, W.S. 1977. Differential sensitivity for smell: noise at the nose. Science, 195, 795–798.
Cornsweet, T.M. 1962. The staircase method in psychophysics. American Journal of Psychology, 75, 485–491.
Gescheider, G.A. 1997. Psychophysics. The Fundamentals. Third edition. Lawrence Erlbaum, Mahwah, NJ.
Harwood, M.L., Ziegler, G.R., and Hayes, J.E. 2012. Rejection threshold in chocolate milk: evidence for segmentation. Food Quality and Preference, 26, 128–133.
James, W. 1892. Psychology, Briefer Course. Holt, Rinehart and Winston, New York, NY.
Lawless, H.T. 2010. A simple alternative analysis for threshold data determined by ascending forced-choice method of limits. Journal of Sensory Studies, 25, 332–346.
Lawless, H.T. and Heymann, H. 2010. Sensory Evaluation of Food. Springer Publishing, New York, NY.
Lawless, H.T., Thomas, C.J.C., and Johnston, M. 1995. Variation in odor thresholds for L-carvone and cineole and correlations with suprathreshold intensity ratings. Chemical Senses, 20, 9–17.
Linschoten, M.R., Harvey, L.O., Eller, P.A., and Jafek, B.W. 1996. Rapid and accurate measurement of taste and smell thresholds using an adaptive maximum-likelihood staircase procedure. Chemical Senses, 21, 633–634.
Morrison, D.G. 1978. A probability model for forced binary choices. American Statistician, 32, 23–25.
Peng, M., Jaeger, S.R., and Hautus, M.J. 2012. Determining odour detection thresholds: Incorporating a method-independent definition into the implementation of ASTM E679. Food Quality and Preference, 25, 95–104.
Prescott, J., Norris, L., Kunst, M., and Kim, S. 2005. Estimating a "consumer rejection threshold" for cork taint in white wine. Food Quality and Preference, 18, 345–349.
Stevens, J.C. 1971. Psychophysics. In: Stimulus and Sensation, Readings in Sensory Psychology. W.S. Cain and L.E. Marks (Eds). Little, Brown and Co., Boston, MA, pp. 5–18.
Stocking, A.J., Suffet, I.H., McGuire, M.J., and Kavanaugh, M.C. 2001. Implications of an MTBE odor study for setting drinking water standards. Journal of the American Water Works Association, 93, 95–105.
Teghtsoonian, R. 1971. On the exponents in Steven's law and the constant in Ekman's law. Psychological Review, 78, 71–80.
USEPA. 2001. Statistical analysis of MTBE odor detection thresholds in drinking water. National Service Center for Environmental Publications (NSCEP) No. 815R01024, available from http://nepis.epa.gov.
Walker, J.C., Hall, S.B., Walker, D.B., Kendall-Reed, M.S., Hood, A.F., and Nio, X.-F. 2003. Human odor detectability: new methodology used to determine threshold and variation. Chemical Senses, 28, 817–826.