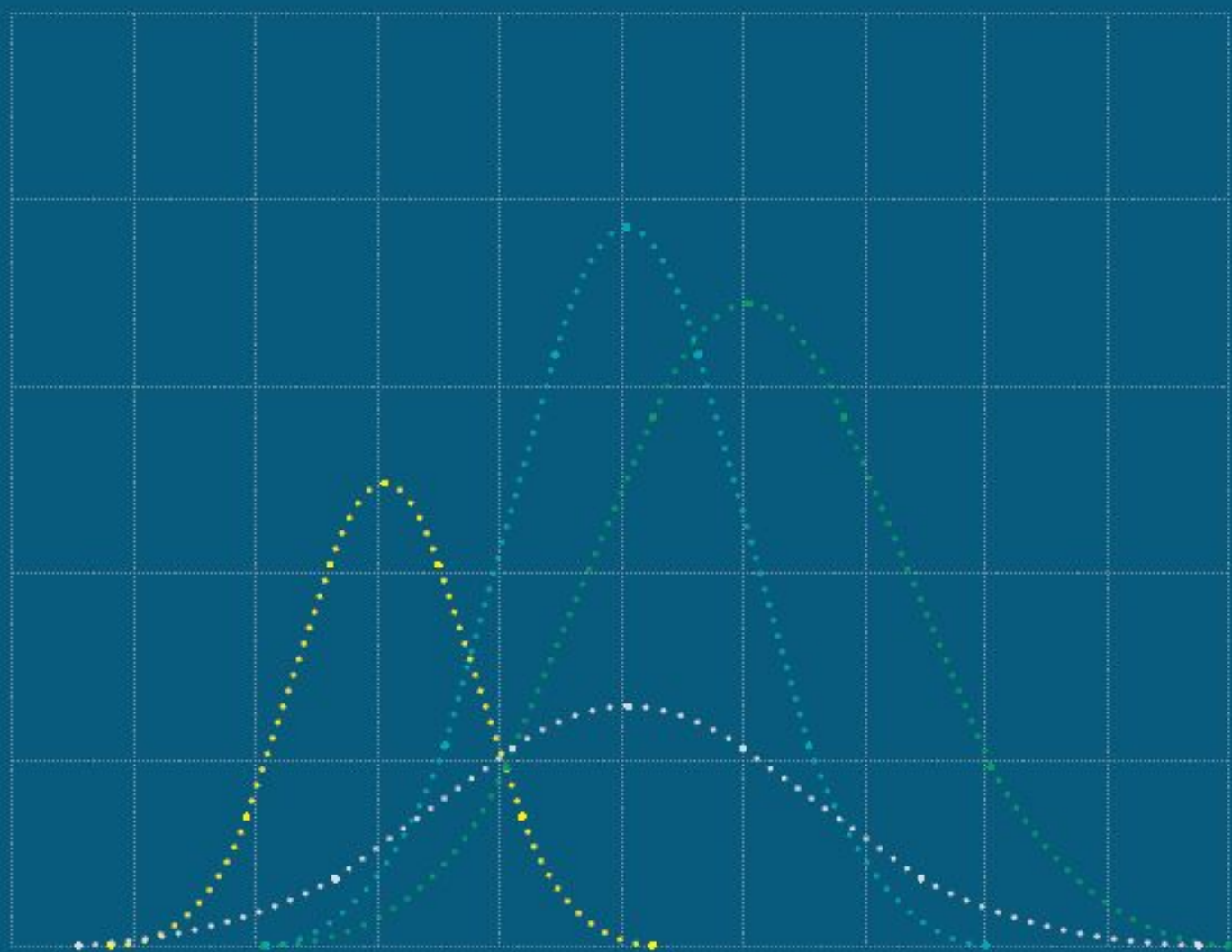


Quantitative Sensory Analysis

Psychophysics, Models and Intelligent Design

Harry T. Lawless



WILEY Blackwell

Contents

[Preface](#)

[1 Psychophysics I: Introduction and Thresholds](#)

[1.1 Introduction and Terminology](#)

[1.2 Absolute Sensitivity](#)

[1.3 Methods for Measuring Absolute Thresholds](#)

[1.4 Differential Sensitivity](#)

[1.5 A Look Ahead: Fechner's Contribution](#)

[Appendix 1.A Relationship of Proportions, Areas Under the Normal Distribution, and Z-Scores](#)

[Appendix 1.B Worked Example: Fitting a Logistic Function to Threshold Data](#)

[References](#)

[2 Psychophysics II: Scaling and Psychophysical Functions](#)

[2.1 Introduction](#)

[2.2 History: Cramer, Bernoulli, Weber, and Fechner](#)

[2.3 Partition Scales and Categories](#)

[2.4 Magnitude Estimation and the Power Law](#)

[2.5 Cross-Modality Matching; Attempts at Validation](#)

[2.6 Two-Stage Models and Judgment Processes](#)

[2.7 Empirical Versus Theory-Based Functions](#)

[2.8 Hybrid Scales and Indirect Scales: A Look Ahead](#)

[2.9 Summary and Conclusions](#)

[Appendix 2.A Decibels and Sones](#)

[Appendix 2.B Worked Example: Transformations Applied to Non-Modulus Magnitude Estimation Data](#)

[References](#)

[3 Basics of Signal Detection Theory](#)

[3.1 Introduction](#)

[3.2 The Yes/No Experiment](#)

[3.3 Connecting the Design to Theory](#)

[3.4 The ROC Curve](#)

[3.5 ROC Curves from Rating Scales; the R-Index](#)

[3.6 Conclusions and Implications for Sensory Testing](#)

[Appendix 3.A Table of \$p\$ and \$Z\$](#)

[Appendix 3.B Test for the Significance of Differences Between \$d'\$ Values](#)

[References](#)

[4 Thurstonian Models for Discrimination and Preference](#)

[4.1 The Simple Paired-Choice Model](#)

[4.2 Extension into n-AFC: The Byer and Abrams "Paradox"](#)

[4.3 A Breakthrough: Power Analysis and Sample Size Determination](#)

[4.4 Tau Versus Beta Criteria: The Same-Different Test](#)

[4.5 Extension to Preference and Nonforced Preference](#)

[4.6 Limitations and Issues in Thurstonian Modeling](#)

[4.7 Summary and Conclusions](#)

[Appendix 4.A The Bradley-Terry-Luce Model: An Alternative to Thurstone](#)

[Appendix 4.B Tables for delta Values from Proportion Correct](#)

[References](#)

[5 Progress in Discrimination Testing](#)

[5.1 Introduction](#)

[5.2 Metrics for Degree of Difference](#)

[5.3 Replication in Choice Tests](#)

[5.4 Current Variations](#)

[5.5 Summary and Conclusions](#)

[Appendix 5.A Psychometric Function for the Dual Pair Test, Power Equations, and Sample Size](#)

[Appendix 5.B Fun with \$\gamma\$](#)

[References](#)

[6 Similarity and Equivalence Testing](#)

[6.1 Introduction: Issues in Type II Error](#)

[6.2 Commonsense Approaches to Equivalence](#)

[6.3 Allowable Differences and Effect Size](#)

[6.4 Further Significance Testing](#)

[6.5 Summary and Conclusions](#)

References

7 Progress in Scaling

7.1 Introduction

7.2 Labeled Magnitude Scales for Intensity

7.3 Adjustable and Relative Scales

7.4 Explicit Anchoring

7.5 Post Hoc Adjustments

7.6 Summary and Conclusions

Appendix 7.A Examples of Individual Rescaling for Magnitude Estimation

References

8 Progress in Affective Testing: Preference/Choice and Hedonic Scaling

8.1 Introduction

8.2 Preference Testing Options

8.3 Replication

8.4 Alternative Models: Ferris k-vsit, Dirichlet Multinomial

8.5 Affective Scales

8.6 Ranking and Partial Ranking

8.7 Conclusions

Appendix 8.A Proof that the McNemar Test is Equivalent to the Binomial Approximation Z-Test (AKA Sign Test)

References

9 Using Subjects as Their Own Controls

Part I: Designs using Parametric Statistics

9.1 Introduction to Part I

9.2 Dependent Versus Independent t-Tests

9.3 Within-Subjects ANOVA (“Repeated Measures”)

9.4 Issues

Part II: Nonparametric Statistics

9.5 Introduction to Part II

9.6 Applications of the McNemar Test: A-not-A and Same-Different Methods

9.7 Examples of the Stuart-Maxwell

9.8 Further Extensions of the Stuart Test Comparisons

9.9 Summary and Conclusions

Appendix 9.A R Code for the Stuart Test

References

10 Frequency Counts and Check-All-That-Apply (CATA)

10.1 Frequency Count Data: Situations — Open Ends, CATA

10.2 Simple Data Handling

10.3 Repeated or Within-Subjects Designs

10.4 Multivariate Analyses

10.5 Difference from Ideal and Penalty Analysis

10.6 Frequency Counts in Advertising Claims

10.7 Conclusions

Appendix 10.A Proof Showing Equivalence of Binomial Approximation Z-Test and χ^2 Test for Differences of Proportions

References

11 Time-Intensity Modeling

11.1 Introduction: Goals and Applications

11.2 Parameters Versus Average Curves

11.3 Other Methods and Analyses

11.4 Summary and Conclusions

References

12 Product Stability and Shelf-Life Measurement

12.1 Introduction

12.2 Strategies, Measurements, and Choices

12.3 Study Designs

12.4 Hazard Functions and Failure Distributions

12.5 Reaction Rates and Kinetic Modeling

12.6 Summary and Conclusions

References

13 Product Optimization, Just-About-Right (JAR) Scales, and Ideal Profiling

13.1 Introduction

13.2 Basic Equations, Designed Experiments, and Response Surfaces

13.3 Just-About-Right Scales

13.4 Ideal Profiling

13.5 Summary and Conclusions

References

14 Perceptual Mapping, Multivariate Tools, and Graph Theory

[14.1 Introduction](#)
[14.2 Common Multivariate Methods](#)
[14.3 Shortcuts for Data Collection: Sorting and Projective Mapping](#)
[14.4 Preference Mapping Revisited](#)
[14.5 Cautions and Concerns](#)
[14.6 Introduction to Graph Theory](#)
[References](#)

[15 Segmentation](#)

[15.1 Introduction](#)
[15.2 Case Studies](#)
[15.3 Cluster Analysis](#)
[15.4 Other Analyses and Methods](#)
[15.5 Women, Fire, and Dangerous Things](#)
[References](#)

[16 An Introduction to Bayesian Analysis](#)

[16.1 Some Binomial-Based Examples](#)
[16.2 General Bayesian Models](#)
[16.3 Bayesian Inference Using Beta Distributions for Preference Tests](#)
[16.4 Proportions of Discriminators](#)
[16.5 Modeling Forced-Choice Discrimination Tests](#)
[16.6 Replicated Discrimination Tests](#)
[16.7 Bayesian Networks](#)
[16.8 Conclusions](#)
[References](#)

[Appendix A: Overview of Sensory Evaluation](#)

[A.1 Introduction](#)

[A.2 Discrimination and Simple Difference Tests](#)

[A.3 Descriptive Analysis](#)

[A.4 Affective Tests](#)

[A.5 Summary and Conclusions](#)

[References](#)

[Appendix B: Overview of Experimental Design](#)

[B.1 General Considerations](#)

[B.2 Factorial Designs](#)

[B.3 Fractional Factorials and Screening](#)

[B.4 Central Composite and Box-Behnken Designs](#)

[B.5 Mixture Designs](#)

[B.6 Summary and Conclusions](#)

[References](#)

[Appendix C: Glossary](#)

[Index](#)

To Michael and Patrick

Quantitative Sensory Analysis

Psychophysics, Models and Intelligent Design

Harry T. Lawless, Ph.D.

Professor Emeritus, Cornell University

WILEY Blackwell

This edition first published 2013 © 2013 by John Wiley & Sons, Ltd

Wiley-Blackwell is an imprint of John Wiley & Sons, formed by the merger of Wiley's global Scientific, Technical and Medical business with Blackwell Publishing.

Registered Office

John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Offices

9600 Garsington Road, Oxford, OX4 2DQ, UK

The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

111 River Street, Hoboken, NJ 07030-5774, USA

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

The right of the author to be identified as the author of this work has been asserted in accordance with the UK Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty: While the publisher and author(s) have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. It is sold on the understanding that the publisher is not engaged in rendering professional services and neither the publisher nor the author shall be liable for damages arising herefrom. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Lawless, Harry T.

Quantitative sensory analysis / Harry T. Lawless.

pages cm.

Includes bibliographical references and index.

ISBN 978-0-470-67346-1 (cloth)

1. Food-Sensory evaluation. 2. Chemistry, Analytic-Quantitative. I. Title.

TX546.L378 2014

664'.07-dc23

2013008677

A catalogue record for this book is available from the British Library.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Cover design and illustration by Sophie Ford
www.hisandhersdesign.co.uk

Preface

It was my intention that the book might serve as a text or companion for an upper level course in sensory evaluation, perhaps as a second semester offering following the introductory course in sensory evaluation. As such, it is aimed at beginning graduate students, advanced undergraduates, and practitioners with a solid background in theory. That being said, it is nearly impossible to write a book that is truly state of the art because the field is so rapidly developing and new models and theories abound each year. This is especially true in sensometrics. Rather than dwell on the latest popular theory or development, I have focused on fundamentals, mathematical principles, and models that I suspect will stand the test of time.

This book makes some assumptions about the experience and background of the reader. It is assumed that the reader has some familiarity with applied sensory evaluation methods as they are used in the foods and consumer products industries. Therefore, the book spends a minimal amount of space defining the methods, and it avoids elaborate descriptions of all the variations, pros and cons, pitfalls, and controversies of sensory testing procedures. For further information about the practical aspects of sensory testing, many books are available, notably *Sensory Evaluation of Foods, Principles and Practices* by Lawless and Heymann (2010), *Sensory Evaluation Practices* by Stone et al., (2006), and *Sensory Evaluation Techniques* by Meilgaard et al. (2006). These three are lengthy textbooks. Shorter guides for the novice reader are include the ASTM document *Sensory Testing Methods* by Chambers and Wolf (1996) and the practical guide *Sensory Evaluation, A Practical*

Handbook by Kemp et al. (2009). A brief overview of sensory techniques is provided in Appendix A.

I have also assumed that the reader has some passing familiarity with the foods or consumer products industries, and possibly the flavor and fragrance industry. This is not meant to be a text on psychophysics per se, so the areas of visual and auditory sensation and perception are rarely mentioned. A fundamental background in basic chemistry and biology is also assumed. Mathematically, I have tried to avoid calculus and matrix algebra as much as possible, so that a reader with basic algebra skills can understand the models and equations, their variations and transformations. It is difficult to know the level of detail that different readers will see as mathematically appropriate or challenging. So the information and models herein are those I deem necessary to understand for the well-trained sensory scientist operating in the food, beverage, or consumer products industries or in an academic setting. To that extent, the book is intended to form a basis or starting point for further study of individual issues and models. Worked examples are provided in most sections to illustrate the application of the equations that have been presented.

The obvious technical area of sensory evaluation that is neglected in this book is the realm of descriptive analysis. I chose not to deal with issues in panel leadership and terminology development and take only a quick look at panelist monitoring. The first topic is primarily qualitative and deals with techniques and problems in human interaction. It is best taught by going to a workshop, participating in panels with a good leader, and then leading a panel yourself. The old rule for learning an operation or procedure among surgical residents was “watch one, do one, teach one.” Perhaps that rule applies in being a descriptive panel leader as well. Terminology development is primarily conceptual, qualitative, and language driven,

and as such is not truly a quantitative technique. Panelist monitoring is primarily a statistical issue. However, descriptive techniques are touched upon in the sections on scaling and panelist calibration.

Another controversial topic was multivariate analyses and perceptual mapping. There has been an explosion in techniques used for multivariate analysis of sensory data, and most of these result in the production of a biplot, a two-dimensional representation of sensory space, with products and assessors as points and with attributes as vectors. Probably more than half the methodological papers in journals like *Food Quality and Preference* pertain to these techniques. Certainly, it is a central theme of sensometrics. There is nothing wrong with this data reduction and mapping approach, and it can be most valuable in looking at competitive monitoring (where are my products relative to the competition?) in situations like product category appraisals. Yet, I remain somewhat skeptical as to how much these techniques can help a sensory practitioner on a day-to-day basis. Most of the normal and often tedious sensory testing that goes on in a large food company concerns issues like cost reduction, ingredient substitution, improvements in the nutritional profile of a product, producing successful variations on a current product, quality control, and shelf-life testing. Very few of these endeavors require any multivariate techniques. So, they are dealt with in the topic of product optimization, and one later chapter gives a general overview of perceptual mapping, which necessarily entails some multivariate techniques.

It should also be clear to the reader that I did not intend to write a sensory statistics guide. Good resources are available, such as the appendices to Lawless and Heyman (2010), the excellent short treatise on analysis of variance by Lea et al. (1998), the sensory statistics text by Naes et al. (2010), and the book on discrimination testing by Jian Bi

(2006). An older resource is the sensory statistics book by O'Mahony (1986), which contains much pithy advice. Some less well known techniques are presented, however, where I felt the penetration of these tests into the mainstream was insufficient. An example is the use of the Stuart-Maxwell test for analysis of categorical or nominal data in a simple two-product test and a complete block design (analogous to the paired or dependent t -test for scaled data).

The structure of the book proceeds as follows. The first four chapters are foundational, dealing with psychophysics. The next four deal with basic topics from a sensory evaluation perspective, namely discrimination, sensory equivalence, scaling, and hedonic testing. I have tried to incorporate current theoretical and practical developments, so their titles include the word "progress." Chapter 9 is an overview of some principles in experimental design that I refer to as "intelligent." The next three chapters cover some applied topics, including check-all-that-apply and other frequency count data, time-intensity modeling, and shelf-life testing. More advanced topics come next, including product optimization methods, perceptual mapping, and consumer segmentation. Note that some of these use multivariate techniques; however, I chose not to include a specific chapter on multivariates (there are several useful books) but rather to focus on how they are applied. At the end, I have provided a forward-looking chapter on the topic of Bayesian analysis, because I felt a graduate student in sensory evaluation should at least understand the principles involved. Appendices are provided, first as a primer in sensory evaluation, second to show some common experimental designs not mentioned elsewhere, and third to provide a glossary.

Many mentors and collaborators have helped, directly or indirectly with the construction of this work. The inspiration for my career in quantitative sensory work starts with a

course in sensory psychology that was team-taught in the 1970s by the scientists at the John B. Pierce Foundation, notably (in alphabetical order) Ellie Adair, Linda Bartoshuk, Bill Cain, Larry Marks, and Joe Stevens. I owe a special debt to Linda Bartoshuk for teaching me that psychophysics could provide a window into sensory physiology and to Bill Cain for providing so many good examples of the parametric quantification of sensory function.

Other mentors in psychophysics have been major influences during my career, notably Trygg Engen, Don McBurney, David Stevens, Michael O'Mahony, Armand Cardello, Howard Schutz, Herb Meiselman, and Howard Moskowitz. This book is aimed at workers in food science, and so a debt must be acknowledged to the late Rose Marie Pangborn, to David Peryam, and to Hildegard Heymann. The writings of Daniel Ennis and colleagues, and also of Jian Bi, were highly influential in many sections of this book, as seen in the citations and reference lists. I thank them for bringing the field to a higher level. The Institute for Perception is thanked for providing the valuable compendium "Short Stories in Sensory and Consumer Science."

Specific assistance was received from many sources. George Gescheider was a helpful correspondent during the writing phase, and his book on psychophysics was a major resource. I would like to give a special note of appreciation to Dr. Guillermo Hough for producing his excellent little book on sensory shelf life, and conducting workshops. I was fortunate to take his workshop on shelf-life methods in Florence in 2009. Much of Chapter 12 is derived from his book and workshop course notes. My former student Michael Nestrud was instrumental in providing the outline for the section on graph theory. Nort Holschuh of General Mills suggested the scoring scheme for the same-different test with a sureness rating scale. Kevin Blot and Rui Xiong

from Unilever assisted in analysis of simulation data. Benoit Rousseau provided valuable discussions concerning Thurstonian analysis. Thierry Worch and Pieter Punter provided useful literature for the section on ideal profiling.

Humans are natural comparators. We are fundamentally change-detectors. The status quo is rarely of interest. The smell of the other apes in the cave is not important. The sudden smell of the saber-toothed cat at the mouth of the cave is. Sensory test methods that take into account this human ability and use it appropriately are destined to be useful tools.

Harry T. Lawless
Ithaca, New York, 2013

References

Chambers, E.C., IV, and Wolf, M.B. 1996. Sensory Testing Methods. Second edition. ASTM Manual Series MNL 26. ASTM, West Conshohocken, PA.

Kemp, S.E., Hollowood, T., and Hort, J. 2009. Sensory Evaluation. A Practical Handbook. John Wiley & Sons, Ltd, Chichester, UK.

Lawless, H.T. and Heymann, H. 2010. Sensory Evaluation of Foods. Principles and Practices. Second edition. Springer Science + Business, New York, NY.

Lea, P., Naes, T., and Rødbotten, M. 1998. Analysis of Variance for Sensory Data. John Wiley & Sons, Ltd, Chichester, UK.

Meilgaard, M., Civille, G.V., and Carr, B.T. 2006. Sensory Evaluation Techniques. Third edition. CRC Press, Boca Raton, FL.

Naes, T., Brockhoff, P.B., and Tomic, O. 2010. Statistics for Sensory and Consumer Science. John Wiley & Sons, Ltd, Chichester, UK.

O'Mahony, M. 1986. Sensory Evaluation of Food, Statistical Methods and Procedures. Marcel Dekker, New York, NY.

1 Psychophysics I: Introduction and Thresholds

[1.1 Introduction and Terminology](#)

[1.2 Absolute Sensitivity](#)

[1.3 Methods for Measuring Absolute Thresholds](#)

[1.4 Differential Sensitivity](#)

[1.5 A Look Ahead: Fechner's Contribution](#)

[Appendix 1.A: Relationship of Proportions, Areas Under the Normal Distribution, and Z-Scores](#)

[Appendix 1.B: Worked Example: Fitting a Logistic Function to Threshold Data](#)

[References](#)

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 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

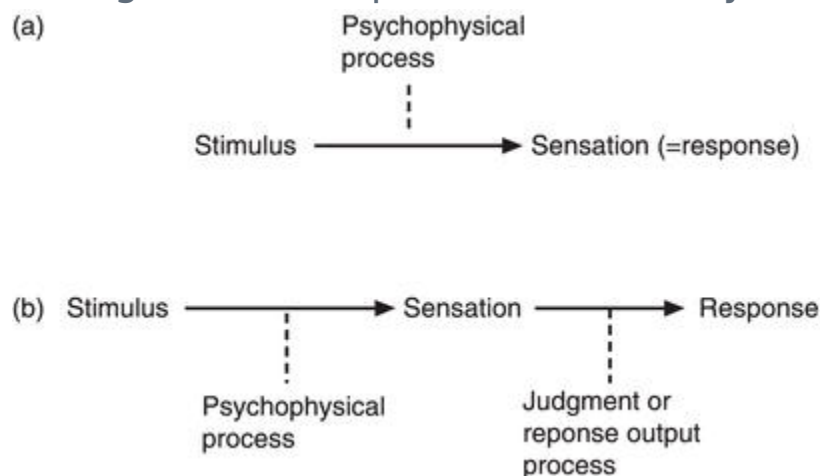
one or the other of the two variables to produce a constant sensation intensity. Thus, the tradeoff function tells us about the ability of the eyes to integrate the energy of very brief stimuli over time. Similar tradeoff functions can be studied for the ability of the auditory system to integrate the duration and sound pressure of a very brief tone in order to produce 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 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.

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.



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 has 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

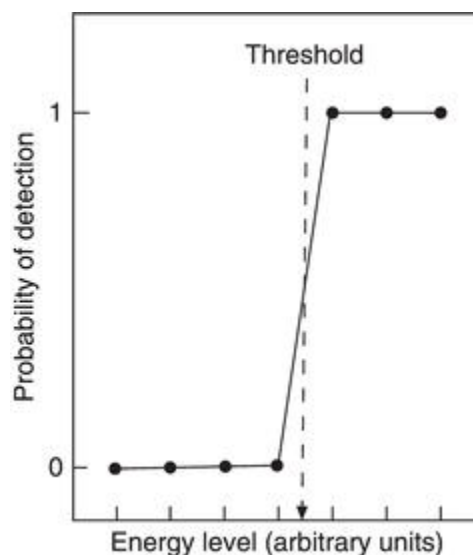
1.2.1 The Threshold Concept

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 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.

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



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 μ) and a spread parameter (the standard deviation σ or variance σ^2). The exact form that generates our familiar bell shaped curve is given by

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

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

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

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 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:

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

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:

$$(1.4) \quad \ln\left(\frac{p}{1-p}\right) = b_0 + b_1x$$

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.

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.

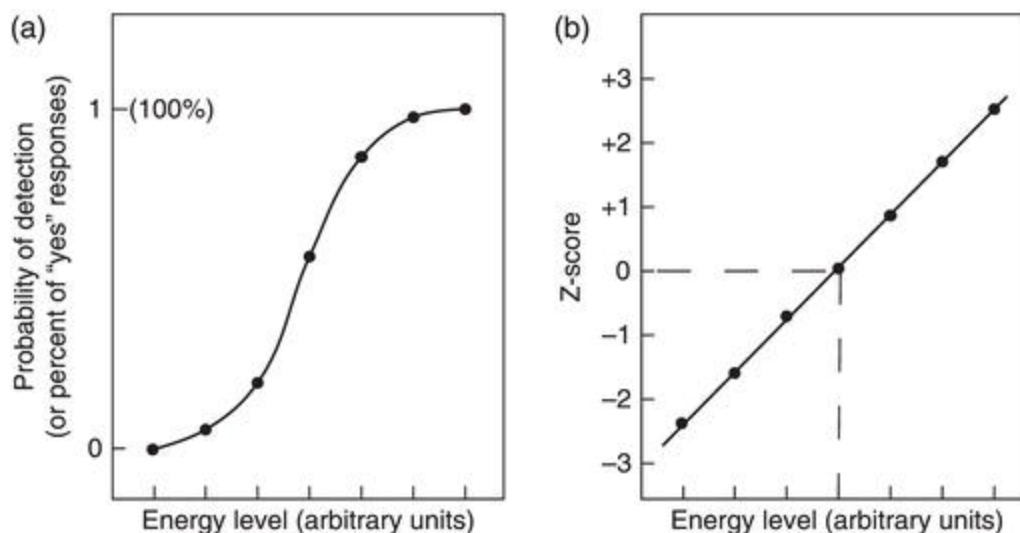


Table 1.1 Conversion of threshold curve proportions to Z-scores

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

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 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-