CHAPTER 3

# Assessing Duration Discrimination: Psychophysical Methods and Psychometric Function Analysis

*Karin M. Bausenhart, Massimiliano Di Luca and Rolf Ulrich*

## 1 Introduction

An important aspect in timing and time perception research is investigating the ability to perceive and compare temporal intervals, that is, the study of duration discrimination (Bindra & Waksberg 1956; Grondin 2010; Matthews & Meck 2016). Just as in every perceptual domain, a central problem in this field is how the relation between physical stimulus input (e.g., a tone lasting for 500 ms) and the sensation evoked by this input (the perceived duration of this stimulus) can be quantified. The scientific study of this relation is called psychophysics (Fechner 1889; Gescheider 1997).

One fundamental issue in psychophysics is the measurement of the difference threshold (just noticeable difference, $JND$; difference limen, $DL$), or in other terms, discrimination sensitivity. It is often loosely defined as the minimal physical difference between two stimuli (e.g., a 500 ms vs. a 550 ms interval) that a participant can just notice. A second important concept in psychophysics concerns the magnitude of the sensation evoked by a given stimulus. Typically, this sensation magnitude is determined by identifying the physical magnitude of a stimulus that is judged to be equal to the magnitude of another stimulus defined as the standard stimulus. For example, one might pinpoint that an auditorily presented temporal interval must be 480 ms to appear as having the same duration as a visually presented standard interval of 500 ms duration. This point along the duration dimension is termed the point of subjective equality ($PSE$), and just as in the example above, it often does not correspond to the point of objective equality ($POE$), which indexes physical equality with the standard stimulus.

Although these definitions appear simple, the experimental determination of these indices of discrimination performance can be quite cumbersome. For example, $PSE$ can be influenced by perceptual and decisional biases, and this may even depend on the specific procedures employed for data collection. For example, when a participant is asked to compare the duration of two

successively presented identical intervals, there may be a general tendency (i.e., bias) to judge the second presented duration as longer than the first presented one. Another major problem is that discrimination performance randomly fluctuates from moment to moment. For example, sometimes a given physical difference between two stimuli is perceived, while sometimes this difference is not perceived. To overcome such obstacles, 19th century researchers already invented various psychophysical tools for measuring discrimination performance, such as the method of constant stimuli (Hegelmaier 1852; Renz & Wolf 1856).

In this chapter, we review several of these tools and methods that are especially useful for measuring duration discrimination performance. Numerical examples are provided to illustrate these psychophysical procedures. In the first section, we introduce the standard psychometric function for comparative judgments and its associated parameters. We discuss various experimental paradigms, which are typically used to collect such data for assessing discrimination performance. In the second section, we present data collection and analysis methods based on equality judgments. For each type of judgment, we introduce several parametric and non-parametric procedures for computing indices of discrimination performance from these data, including exemplary Matlab scripts implementing these procedures (see book's GitHub repository). In the final conclusion, we briefly review several advanced toolboxes available for assessing discrimination performance.

## 2      Comparative Judgments

Several of the experimental paradigms, which are typically employed in timing research, involve comparative judgments. Specifically, these judgments require that participants decide whether a given stimulus duration is longer or shorter than a certain target duration. For example, in the so-called reminder task, the participant receives two successive durations in each experimental trial. One of the two durations is the target duration that is kept constant across a block of trials. This duration is traditionally called the *standard* or reference duration $s$ (Guilford 1954; Woodworth & Schlosberg 1954). The other duration varies randomly from trial to trial and is usually called the *comparison* or test duration $c$.

In most experiments, several different comparison durations are used, some larger than $s$ and some smaller than $s$. Typically, between 6 and 12 different values of $c$ are arranged symmetrically around $s$. It is convenient to index these

comparison levels as $c_1,\ldots,c_k$ from the smallest to the largest.[1] These selected comparison levels are presented several times (usually 10 to 20 repetitions) during the course of a single experiment in a random sequence. The order of $s$ and $c$ may be either constant (fixed stimulus order), for example, in each trial $s$ is presented first, or it may vary randomly from trial to trial (random stimulus order). In the following, we will introduce some typical experimental paradigms employing either fixed or random stimulus order and describe several methods for analyzing the data emerging from these paradigms.

## 2.1 Fixed Order of Standard and Comparison Stimuli

Presumably the most elementary psychophysical approach uses a fixed order of $s$ and $c$ (e.g., Luce & Galanter 1963). For example, in the classical *reminder task*, $s$ precedes $c$ in every trial. Participants are typically asked whether the first or second stimulus appears longer, and consequently select the response $R_1$ or $R_2$, respectively. It is important to note that participants have to choose one of the two response alternatives in every trial – if a judgment cannot be made with certainty, the subject is asked to choose the alternative that seems most appropriate or simply to guess an alternative. After each trial, the experimenter simply records whether the participant responded with $R_1$ or $R_2$.

Table 3.1 contains an outcome example of such a psychophysical experiment comprising $k = 9$ comparison durations centered symmetrically around $s = 500$ ms. For these data, the relative frequency $f_i$ of responding with $R_2$ as a

TABLE 3.1    *Exemplary outcome of a psychophysical experiment with* s = 500 ms *and nine comparison levels $c_1,\ldots,c_9$ ranging from $c_1$ = 300 ms to $c_9$ = 700 ms. Row $n_i$ shows how many responses per comparison level $c_i$ were recorded during the course of the experiment. For most c levels, data of 15 trials were available, yet for some levels fewer data were recorded – for example, because the participant forgot to respond or occasionally pressed the wrong response key on the keyboard. The rows $n_{1,i}$ and $n_{2,i}$ show the number of $R_1$ and $R_2$ responses, respectively. The row $f_i$ contains the relative frequency of $R_2$ responses per c level.*

|           | $c_1$ | $c_2$ | $c_3$ | $c_4$ | $c_5$ | $c_6$ | $c_7$ | $c_8$ | $c_9$ |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|           | 300   | 350   | 400   | 450   | 500   | 550   | 600   | 650   | 700   |
| $n_i$     | 15    | 15    | 14    | 15    | 12    | 15    | 15    | 15    | 14    |
| $n_{1,i}$ | 14    | 11    | 12    | 5     | 3     | 0     | 1     | 0     | 0     |
| $n_{2,i}$ | 1     | 4     | 2     | 10    | 9     | 15    | 14    | 15    | 14    |
| $f_i$     | 0.07  | 0.27  | 0.14  | 0.67  | 0.75  | 1.00  | 0.93  | 1.00  | 1.00  |

---

1    Usually the two extreme values in the range, $c_1$ and $c_k$, are selected in such a way that the comparisons cover the full range of the psychometric function from 0 to 1. Weber fractions
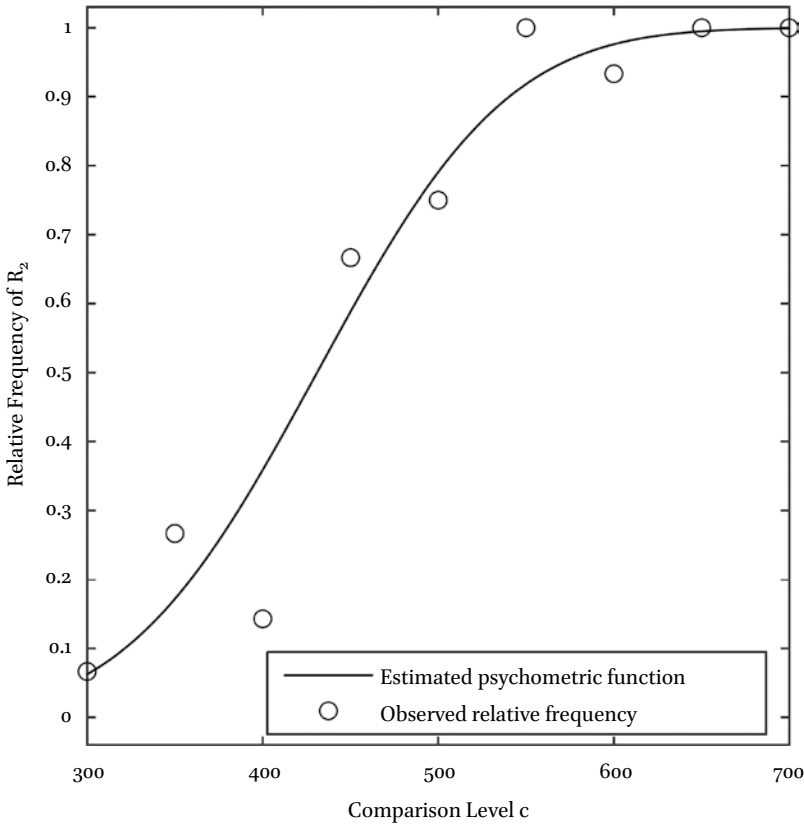
FIGURE 3.1  *Relative frequency of responding with $R_2$ (i.e., judging the second presented duration c as longer than the first presented duration s) as a function of comparison duration (open circles), for the example data given in Table 3.1. The solid line shows the best fitting psychometric function derived by means of probit analysis.*

function of comparison level $c_i$ is depicted in Figure 3.1. Apart from the statistical noise involved in such data, one would expect that this relative frequency increases with increasing duration of $c$.

### 2.1.1      Probit Analysis

In order to enable a more comprehensive analysis of the data emerging from such an experiment, one typically fits a psychometric function $\Psi(c)$ to the relative frequencies of $R_2$ responses per $c$ level (e.g., Luce & Galanter 1963).

---

may help to select these values. For example, assume that $s$ = 500 ms and the participant is asked to discriminate auditory intervals, for which the Weber fraction typically amounts to approximately 0.1 (Rammsayer 2010; Rammsayer & Ulrich 2012). As a rule of thumb, $c_1$ may be selected as $s \cdot (1 - 4 \cdot 0.1)$ and $c_k$ as $s \cdot (1 + 4 \cdot 0.1)$. For $s$ = 500 ms, this would yield $c_1$ = 300 ms and $c_k$ = 700 ms.

For example, the cumulative density function (cdf) of a normal distribution has been often used as a mathematical model for $\Psi(c)$. This function increases monotonically from 0 to 1 with increasing values of $c$ and can be expressed as

$$\Psi(c) = \Phi\left(\frac{c - \mu}{\sigma}\right),\qquad(1)$$

where $\Phi$ denotes the cdf of a standard normal distribution, $\mu$ is the location parameter, and $\sigma$ represents the slope of $\Psi$. This approach of modeling the psychometric function is also called probit analysis (Finney 1952).[2]

The parameter $\mu$ denotes the level of $c$ at which the probability of responding with $R_2$ is equal to 0.5, that is, at this level the two responses $R_1$ and $R_2$ are equally likely. This level is often called the *PSE*, because it denotes the duration of $c$, which is judged to have the same duration as $s$. The *PSE* needs not to be equal to $s$. For example, the *PSE* is often smaller than $s$ because participants usually tend to overestimate the second duration compared to the first one, a phenomenon termed the time-order error (Eisler, Eisler, & Hellström, 2008; Köhler 1923). In general, the difference between objective physical equality and subjective equality has been termed *constant error* (*CE*) and has been defined as *CE* = *PSE* – *s* in the psychophysical literature. Shifts of the *PSE* away from the *POE* may reflect a perceptual or a decisional bias.

A second parameter of major importance that can be computed from a psychometric function is the *DL* or *JND*. This parameter indexes the discrimination sensitivity of a participant, with smaller values of *DL* indicating a higher level of sensitivity. The *DL* is related to the steepness of the psychometric function. It is typically defined as half its interquartile range, that is, $DL = \left(c_{0.75} - c_{0.25}\right)/2$, where $c_{0.75}$ and $c_{0.25}$ represent the stimulus levels at which the response $R_2$ is elicited with probability 0.75 and 0.25, respectively (Luce & Galanter 1963). Consequently, *DL* indexes the duration difference between $s$ and $c$, which enables the subject to identify $c$ as being either shorter or longer than $s$ with an accuracy level of 75%. For the function embodied in Equation 1, the *DL* is given by

$$DL = \sigma \cdot z_{0.75}\qquad(2)$$

---

2   Other functional families than the normal distribution are often used to model the psychometric function, such as the logistic or the Weibull function. However, the logistic and the probit model produce virtually the same results (Lord, Novick, & Birnbaum, 1968, p. 399).

where $z_{0.75}$ is the 75% percentile of the standard normal distribution, i.e., $z_{0.75} \approx 0.6745$.[3]

An especially efficient method for estimating the parameters $PSE$ and $DL$ is Fisher's maximum-likelihood procedure. In brief, one uses Equation 1 to compute the likelihood of the observed data,

$$L\left(Data \mid \mu, \sigma\right) = \Pi_{i=1}^{k} \Psi\left(c_i\right)^{n_{2,i}} \times \left[1 - \Psi\left(c_i\right)\right]^{n_{1,i}} \tag{3}$$

where $n_{1,i}$ and $n_{2,i}$ denote the frequencies of observed $R_1$ and $R_2$ responses at each comparison level (compare Table 3.1). The maximum likelihood estimates of $\mu$ and $\sigma$ are those numerical values that maximize this likelihood function. The maximum of this function can be found numerically using a computer, a procedure known as numerical optimization.

A simple Matlab (R2016b) script ("MLEPsyProbit.m") for performing this optimization is available (see book's GitHub repository). It finds the parameters $\mu$ and $\sigma$ at which the function $L\left(Data \mid \mu, \sigma\right)$ has its extremum. This script requires as input the vectors $\vec{c} = \left(c_1, \ldots, c_k\right)$, $\vec{n}_1 = \left(n_{1,1}, \ldots, n_{1,k}\right)$, and $\vec{n}_2 = \left(n_{2,1}, \ldots, n_{2,k}\right)$ and provides the maximum-likelihood estimates of $PSE$ and $DL$ together with their standard errors and their corresponding 95% confidence intervals as outputs. This script computes the standard errors from the observed Fisher information. Applying the script to the data in Table 3.1, one obtains $PSE = 430.9$ ms, $Se = 13.3$ ms with a 95%-confidence interval of $CI = \left[404.9, 457.0\right]$, and $DL = 57.6$ ms, $Se = 8.8$ ms with $CI = \left[40.5, 74.8\right]$. On the basis of the $PSE$ result, the script computes $CE = -69.1$ ms, $Se = 13.3$ ms with $CI = \left[-95.1, -43.0\right]$. The $CE$ indicates a systematic overestimation of the comparisons relative to the standard duration $s = 500$ ms, which might be attributed, for example, to a negative time-order error. Figure 3.1 depicts the relative

---

3   Several researchers (Treutwein 1995; Treutwein & Strasburger 1999; Wichmann & Hill 2001) have suggested to include also lapse parameters in the estimation of psychometric functions to account for trials in which the participant commits stimulus-independent lapses due to phasic inattention or "finger errors". These events will result in scaled psychometric functions, which do not cover the full range from 0 to 1. Even though such processing failures are rare events, typically estimated to occur in between 0% and 5% of trials (Wichmann & Hill 2001), their presence can nonetheless distort the estimation of $DL$. Therefore, if empirical evidence suggests the presence of lapses, corresponding extended psychometric functions should be used for data analysis (Wichmann & Hill 2001, also see Table 3.3 for a list of tools available for performing such advanced analyses). Models comparison statistics can be used as a principled way of choosing the function with or without lapses.

response proportions $f_i$ from Table 3.1 and the resulting psychometric function derived by this probit analysis, which is a standard psychophysical approach for estimating *PSE*, *DL*, and *CE*.

### 2.1.2    Pseudo-Gaussian Function

Killeen, Fetterman, and Bizo (1997) proposed an alternative to Equation 1 that often provides an excellent fit to observed data (Allan & Gerhardt 2001; Birngruber, Schröter, & Ulrich, 2014; Grondin 2001). This approach takes Weber's law into account, according to which variability in perceived duration should linearly increase with physical duration. Specifically, let *S* and *C* represent the internal representations of the standard *s* and the comparison *c*, respectively. In addition, assume that the internal difference $\Delta = C - S$ follows a normal distribution with mean $E[\Delta|c] = c - (\varepsilon + s)$, where the parameter $\varepsilon$ has the status of a constant error. If the standard deviation of the difference $\Delta$ follows Weber's law $\sigma_c = w \cdot c, w > 0$, then the psychometric function is given by the *Pseudo-Gaussian function*,

$$\Psi\left(c\right) = \Phi\left(\frac{c - \left(\varepsilon + s\right)}{w \cdot c}\right),\tag{4}$$

where $\Phi$ again denotes the cumulative density function of a standard normal variable, and the parameters are the constant error $\varepsilon$ and the Weber fraction *w*.[4] This Pseudo-Gaussian function is actually not a genuine psychometric function because it does not converge to 1. However, this deviation from 1 is negligible for realistic values of *w*. The supplementary Matlab script "MLEPSyPseudoGaussian.m" (see book's GitHub repository) provides maximum likelihood estimates of the parameters $\varepsilon$ and *w*. Applying this script to the data in Table 3.1 yields for $\varepsilon$ an estimate of –86.8 ms, *Se* = 12.3 ms, *CI* = [–110.8, –62.7] and for *w* an estimate of 0.190, *Se* = 0.026, *CI* = [0.138, 0.241].

Moreover, for this Pseudo-Gaussian function, it can be shown that the *PSE* is given by

$$PSE = s + \varepsilon\tag{5}$$

---

4    As a further extension, one may replace $\sigma_c = w \cdot c$ by the generalized Weber's law $\sigma_c = \sqrt{w_1 \cdot c^2 + w_2 \cdot c + w_3}$ (see Killeen et al. 1997). A similar model has been proposed by García-Pérez (2014). Also note that for $w_1 = w_2 = 0$ this extended model becomes a special case of the probit model discussed above.

and the $DL$ by

$$DL = (s + \varepsilon) \cdot \frac{w \cdot z_{0.75}}{1 - \left(w \cdot z_{0.75}\right)^2} \qquad (6)$$

with $z_{0.75} \approx 0.6745$. Inserting the above estimates into these equations yields $PSE = 413.2$ ms and $DL = 53.7$ ms. It can be noticed that these estimates differ numerically from the ones of the standard approach embodied by Equation 1, which must be attributed to the different assumptions underlying both models.

Figure 3.2 depicts the relative response proportions $f_i$ from Table 3.1 and the psychometric function resulting from the Pseudo-Gaussian model. A potential
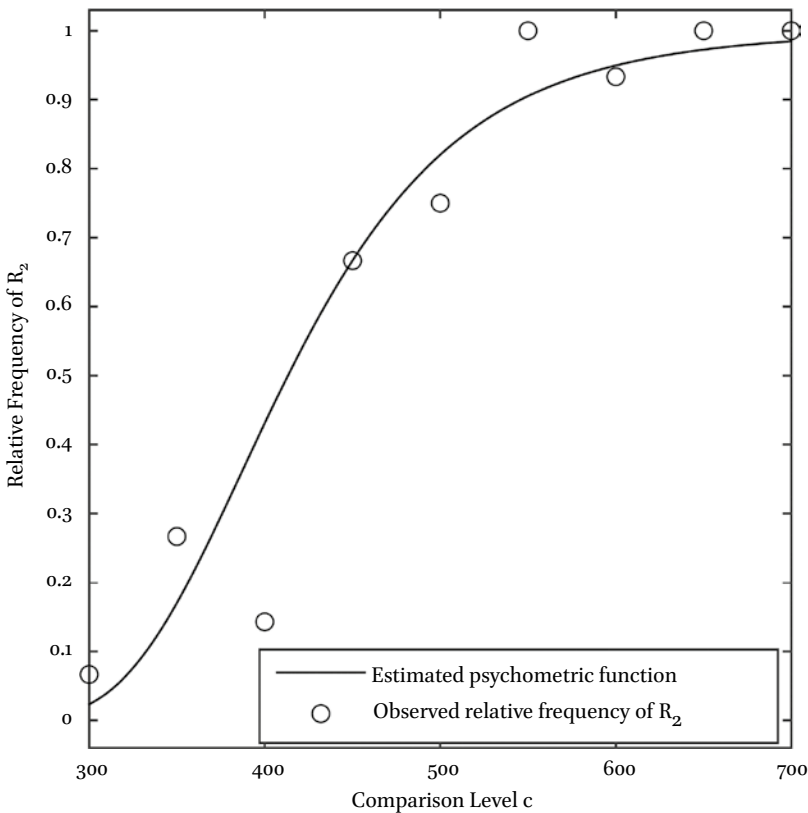


FIGURE 3.2   *Relative frequency of responding with $R_2$ (i.e., judging the second presented duration c as longer than the first presented duration s) as a function of comparison duration (open circles), for the example data given in Table 3.1. The solid line shows the best fitting psychometric function derived by means of the Pseudo-Gaussian model.*

drawback of this Pseudo-Gaussian model is that the estimate of $DL$ is affected by the size of $\varepsilon$, that is, the constant error.

### 2.1.3        Spearman-Kärber Method

In addition to the parametric approaches discussed above, one can also use a nonparametric approach, the *Spearman-Kärber method* (Kärber 1931; Spearman 1908), for estimating the location and the spread of the psychometric function (Miller & Ulrich 2001; Sternberg, Knoll, & Zukofsky, 1982). This method has several advantages. In contrast to parametric approaches, the Spearman-Kärber method does not require specific assumptions about the functional family of the true underlying psychometric functions. Also, it allows for estimating higher-order moments as skewness and kurtosis, in addition to location and spread of the psychometric function. Moreover, this method is computationally efficient compared to others, because it does not require an iterative fitting procedure. Finally, parameter estimates obtained with this method are often even less biased and less variable than parameter estimates obtained by employing parametric approaches (Miller & Ulrich 2001; Ulrich & Miller 2004).

In the Spearman-Kärber method, the range of comparison stimuli is subdivided into bins, each ranging from $c_{i-1}$ to $c_i$, for $i = 1,...,k$. The relative response frequencies $f_i$ associated with each stimulus level $c_i$ are assumed to be uniformly distributed within each corresponding bin. Thus, the probability density within each bin is estimated as $(f_i - f_{i-1})/(c_i - c_{i-1})$. The resulting histogram of probability densities approximates the continuous true cumulative distribution function underlying the data. Each $r$th raw moment $m'_r$ of this psychometric function can then be calculated as

$$m'_r = \frac{1}{r+1} \sum_{i=1}^{k+1} \frac{(f_i - f_{i-1}) \cdot \left(c_i^{r+1} - c_{i-1}^{r+1}\right)}{c_i - c_{i-1}}. \tag{7}$$

It must be noted that in this calculation, the values of the most extreme comparison levels $c_0$ and $c_{k+1}$ are not included in the actual experimental design but must be determined such that true values of $f_0 = 0$ and $f_{k+1} = 1$ can be assumed.

This step is crucial whenever $f_1 > 0$ or $f_k < 1$, that is, whenever the observed psychometric function is truncated (i.e., it does not start at 0 or reach 1). For example, this may be the case if the chosen range of comparison levels for testing was not broad enough to cover the whole range of the psychometric function. Similarly, lapses, finger errors, or simply binomial random error might cause such truncated psychometric functions. In this case, the specific values chosen for $c_0$ and $c_{k+1}$ will affect the CDF's raw moments, and consequentially the

estimates of location, spread, etc. derived from these raw moments. Therefore, it is advised to interpret parameters obtained by means of the Spearman-Kärber method especially carefully whenever truncated psychometric functions are present. Furthermore, for computing higher raw moments (i.e., $r > 1$), it is necessary to monotonize the observed psychometric function before computing these moments with Equation 7 (see Ayer, Brunk, Ewing, Reid, & Silverman, 1955; Miller & Ulrich 2001, cf. also Figure 3.3).

From the raw moments, one can derive estimates of location, spread, skewness and kurtosis (Miller & Ulrich 2001). For example, the first raw moment $m_1'$ corresponds to the arithmetic mean and, thus, indexes the location of the psychometric function (i.e., it serves as an estimate of PSE). The standard deviation of the underlying CDF can be estimated with $\sigma = \sqrt{m_2' - \left(m_1'\right)^2}$.
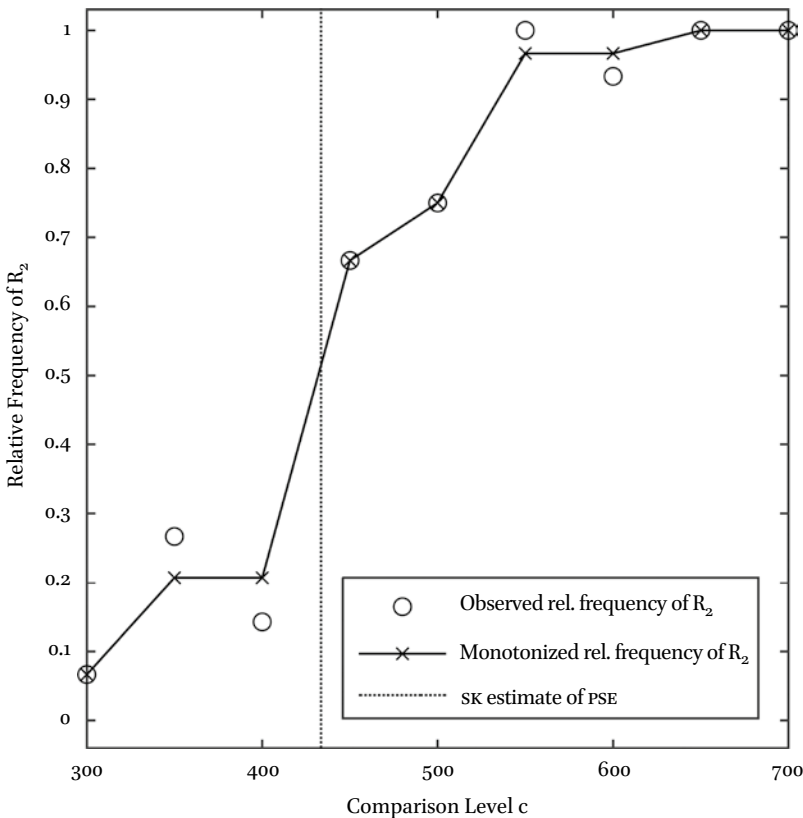


FIGURE 3.3 *Observed (open circles) and monotonized (black X and solid line) relative frequency of responding with $R_2$ (i.e., judging the second presented duration c as longer than the first presented duration s) as a function of comparison duration, for the example data given in Table 3.1. The dotted vertical line corresponds to the PSE estimate derived by the Spearman-Kärber method.*

For comparative purposes, a convenient estimate of $DL$ can then be approximated by multiplying $\sigma$ by $z_{0.75} \approx 0.6745$.

The provided Matlab script "SpearmanKaerber.m" (see book's GitHub repository) monotonizes the observed psychometric function and then computes the Spearman-Kärber estimates of $PSE$, $\sigma$, and $DL$ for the example data contained in Table 3.1 (see Figure 3.3). By default, the extreme values $c_0$ and $c_{k+1}$ are set such that $c_1 - c_0 = c_2 - c_1$, and $c_{k+1} - c_k = c_k - c_{k-1}$, that is, equidistance between the first 3 and the last 3 comparison levels is assumed. The script outputs the observed response frequencies $f_i$ and the monotonized response frequencies $\overline{f_i}$, as well as a vector containing estimates of $PSE$, $\sigma$, $DL$, and $CE$. For the example data given in Table 3.1, the corresponding estimates are $PSE = 433.5$ ms, $\sigma = 82.1$ ms, $DL = 55.4$ ms, and $CE = -66.5$ ms. These parameter estimates correspond quite well with the estimates derived by the probit analysis described above. In addition, this function provides bootstrap estimates of these parameters based on 1000 replications, including standard errors and $CIs$. For example, for $PSE$: $Se = 12.1$ ms, $CI = [408.3, 455.4]$, for $\sigma$: $Se = 10.6$ ms, $CI = [58.4, 100.4]$, for $DL$: $Se = 7.2$ ms, $CI = [39.4, 67.7]$, and for $CE$: $Se = 12.1$ ms, $CI = [-91.7, -44.6]$.[5]

### 2.1.4    Variants of Data Collection

In the preceding sections, it is assumed that in each trial a standard $s$ is presented before the comparison duration $c$ (i.e., reminder task). Especially in the domain of timing research, several variants of this basic task have been proposed (for an overview, see Grondin 2010).

First, in the *single-stimulus method* only the comparison is presented in each trial. The participant then classifies each comparison as either short or long, presumably against an internal standard that is quickly formed from experiencing the comparisons during the course of the experiment (Bausenhart, Bratzke, & Ulrich, 2016; Dyjas, Bausenhart, & Ulrich, 2012; Nachmias 2006; Woodworth & Schlosberg 1954). Sometimes, researchers also present a standard $s$ for several times at the beginning of the experiment, in order to provide a more explicit reference for classifying the duration of each comparison as short or long. In either case, when the proportion of "long" responses is plotted against comparison duration, an ogive psychometric function will emerge. Estimating $PSE$ and $DL$ then can proceed in the same manner as in the standard approach outlined above.

---

5  Naturally, these bootstrapped values will randomly fluctuate with each execution of the provided Matlab function.

Second, a further methodological variant of the standard approach is the *bisection method*. Here, at the beginning of the experiment the shortest (i.e., $c_1$) and the longest (i.e., $c_k$) comparisons are presented several times as anchor stimuli. During the experiment, only comparisons are presented (as in the single-stimulus method) and the participant must classify each comparison as more similar to the short or to the long anchor duration (Allan & Gibbon 1991; Wearden, Rogers, & Thomas, 1997). The data analysis again proceeds as outlined above.

Third, in comparative judgments, researchers may allow for a *third response option* besides $R_1$ and $R_2$, i.e., an "uncertain" or "same" response (Woodworth & Schlosberg 1954, pp. 212–217). Historically, two response categories have been preferred over three response categories in psychophysics (Woodworth & Schlosberg 1954, p. 217). Nevertheless it is sometimes useful to employ three categories for theoretical reasons (e.g., Rammsayer & Ulrich 2001; Ulrich 1987) and more complex models of discrimination performance may be fitted to the data emerging from three-response categories to identify the relevant parameters indicating discrimination performance (García-Pérez 2014; García-Pérez & Alcalá-Quintana 2013).

Finally, all data collection variants as described above may be regarded as instances of the method of constant stimuli, in which the researcher preselects a range of comparison levels and typically presents each comparison level for a predetermined number of repetitions, with all trials presented in random order. This has sometimes been criticized as relatively inefficient, since many points along the psychometric function are sampled with an equal and large number of trials. Yet, some of these points, typically those demarcating threshold values as *PSE* and *DL*, are of especially high interest to the researcher, and an efficient data collection procedure might focus on assessing these points with high precision instead. Since the threshold values are of course not known in advance of testing, but depend on the participants' performance, comparison levels then cannot be specified in advance. Rather, the experimenter's decision about which comparison level should be presented in a given trial must depend on the participant's responses given in previous trials. There is a vast number of data collection schemes and analysis variants for such *adaptive testing* procedures (see Kaernbach 1991; Leek 2001; Treutwein 1995), although some caution is required when applying these procedures (e.g., García-Pérez 1998).

## 2.2     *Random Order of Standard and Comparison Stimuli*
In the methodological variants for data collection described in the preceding section, the temporal order of *s* and *c* is either the same in each experimental

trial, or only $c$ is presented. In contrast, in the so-called *two-alternative forced-choice task* (2AFC, sometimes also two-interval forced-choice task or 2IFC), this order of the standard and comparison varies randomly from trial to trial. Thus, in each trial, stimulus order is either $\langle sc \rangle$ or $\langle cs \rangle$. Participants typically indicate whether the first or second stimulus appears longer by responding with $R_1$ or $R_2$, respectively.[6] Since the order of $s$ and $c$ varies randomly, the range of $c$ levels can be restricted to values $c \geq s$, but it is also possible to employ values ranging from $c_1 < s$ to $c_k > s$. In the latter case, $R_1$ and $R_2$ responses can be recoded as $c > s$ responses, for the stimulus orders $\langle cs \rangle$ and $\langle sc \rangle$, respectively. From these data, a psychometric function depicting the proportion of $c > s$ responses emerges. Given a sufficiently large range of $c$ values, and disregarding the possibility of lapses or finger errors, this function covers the full range from 0 to 1. Then, as a measure of discrimination sensitivity, $DL$ is often estimated as half the interquartile range of this psychometric function (analogously to the procedure outlined above for the reminder task). In the former case, researchers often plot the proportion of correct responses (i.e., $c > s$ responses), resulting in a psychometric function restricted from 0.5 (i.e., guessing probability) to 1. An often-employed procedure to derive $DL$ from such psychometric functions is to compute it as $DL = c_{0.75} - s$ (cf. Ulrich 2010; Ulrich & Vorberg 2009).

In both cases outlined above, however, the common practice of collapsing the raw data across the two orders of $s$ and $c$ can lead to loss of information and even to severe distortions in the estimated parameters of the psychometric function. To avoid such distortions, data from the two stimulus orders $\langle sc \rangle$ or $\langle cs \rangle$ should be plotted and analyzed separately (Ulrich 2010; Ulrich & Vorberg 2009). Consequently, two order-dependent psychometric functions emerge in the 2AFC design (cf. Figure 3.4). Specifically, let $S_1$ and $S_2$ denote the stimulus in the first or second position, respectively. Define $F_1(c) \equiv P(R_1 \mid \langle cs \rangle)$ and $F_2(c) \equiv P(R_2 \mid \langle sc \rangle)$ as the conditional probability with which the participant judges the comparison $c$ as the larger of the two stimuli when it was presented first or second, respectively. Note that the two conditional psychometric functions monotonically increase with $c$.

Importantly, these two conditional psychometric functions can differ in their location ("Type A order effect") and in their spread ("Type B order effect"). A prominent example for a Type A order effect is the typically observed negative time-order error, in which the duration of the first of two subsequently presented intervals is underestimated compared to the second one (Eisler et al. 2008; Hellström 1985; Köhler 1923). Specifically, this would correspond to

---

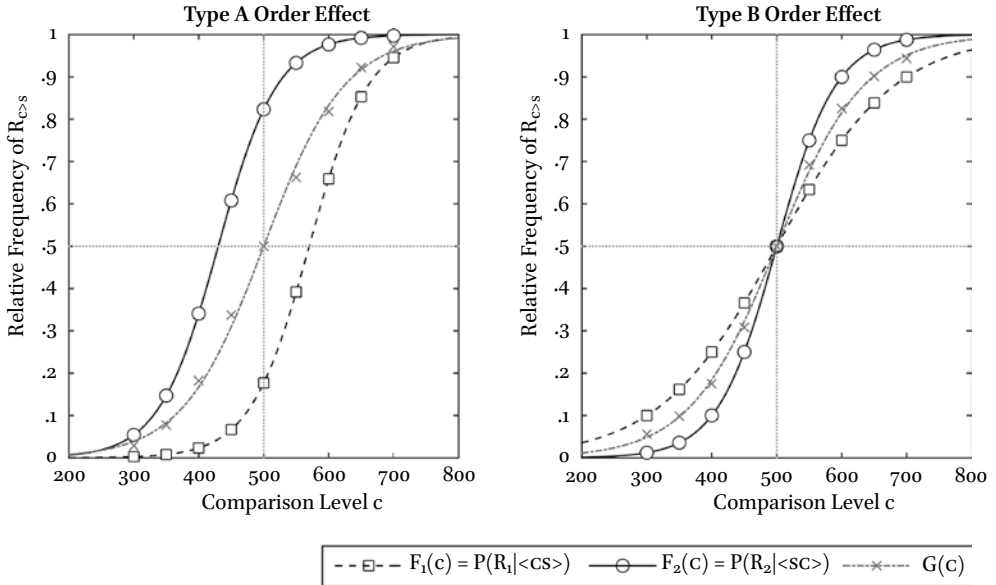6  Participants are usually not aware that there is a constant standard, which appears first or
   second.

FIGURE 3.4 *Relative frequency of responding with $R_{c>s}$ (i.e., judging the comparison duration c as longer than the standard duration s) and psychometric functions for a hypothetical 2AFC experiment. Depicted are the order-conditional functions $F_1(c)$ (dashed line and squares) and $F_2(c)$ (solid line and circles) for stimulus orders ⟨cs⟩ and ⟨sc⟩, respectively. In addition, the psychometric function $G(c)$ (grey dash-dotted line and x) corresponds to the observed response frequencies aggregated across presentation orders. The left panel depicts a type A order effect and the right panel a type B order effect. These effects will be concealed by the common practice of fitting a single psychometric function to the data aggregated across stimulus orders. Moreover, whenever a Type A order effect is present in the data, the aggregated psychometric function $G(c)$ is less steep than either of the order-conditional functions (see left panel). Consequently, DLs derived from such aggregated functions will be overestimated.*

a lateral shift of the conditional psychometric functions away from the *POE*, at which *s* = *c*, such that the location of the conditional psychometric function for stimulus order ⟨cs⟩ is shifted to *POE* + $\gamma$ and the mean location of conditional psychometric function for stimulus order ⟨sc⟩ is shifted to *POE* − $\gamma$. An example of a Type B order effect, which is often observed in duration discrimination, is a shallower slope of the conditional psychometric function for stimulus order ⟨cs⟩ than for stimulus order ⟨sc⟩ (Bruno, Ayhan, & Johnston, 2012; Dyjas et al. 2012; Nachmias 2006). Consequently, such an effect indicates a higher discrimination sensitivity for two subsequent intervals, when the first of these intervals is a standard interval with constant duration, rather than when it is varied randomly from trial to trial.

The preceding explanation assumes that researchers choose the stimuli presented in a 2AFC task such that they vary only along a single stimulus dimension. In a duration discrimination task, for example, $s$ and $c$ would be identical in all respects, except for their duration. In this case, a restriction emerges for the estimated psychometric functions. Specifically, averaging the two conditional functions results in an aggregated psychometric function,

$$G(c) = \frac{P(R_1|\langle cs\rangle) + P(R_2|\langle sc\rangle)}{2}. \tag{8}$$

At the *POE*, defined as $s = c$, this equation simplifies to

$$G(s) = \frac{P(R_1|\langle ss\rangle) + P(R_2|\langle ss\rangle)}{2}. \tag{9}$$

Since $R_1$ and $R_2$ are the only response alternatives, their associated response proportions must sum to one. Consequently,

$$G(s) = \frac{1}{2}, \tag{10}$$

that is, the average of the two order-conditional psychometric functions must pass through the point ($s$, 0.5). This restriction must be considered when fitting psychometric functions to the order-conditional data. Specifically, instead of estimating two independent psychometric functions, they must be fitted simultaneously and the number of the free parameters to be estimated for these two functions reduces to three (Ulrich 2010; Ulrich & Vorberg 2009). Matlab and R code for fitting logistic order-conditional psychometric functions under this restriction is provided by Bausenhart, Dyjas, Vorberg, and Ulrich (2012).

If a researcher chooses to let $s$ and $c$ vary along more than one dimension (e.g., in duration and stimulus size), then of course the constraint implied by Equation 10 does not hold, and the average function will pass through the point (*PSE*, 0.5) instead (García-Pérez & Alcalá-Quintana 2011). Then, the two order-conditional psychometric functions can be estimated independently from each other, just as outlined in the section on fixed order of standard and comparison stimuli above. The routines provided by Bausenhart et al. (2012) also provide the option to release the constraint at $s = c$ and therefore can also be employed for the analysis of order-conditional data coming from 2AFC tasks which vary along multiple stimulus dimensions.

## 3        Equality Judgments

Besides the comparative judgment task employed in the preceding methods, equality judgments as in the *temporal generalization method* are also often used in the domain of temporal cognition (e.g., Wearden 1992; Wearden, Edwards, Fakhri, & Percival, 1998). In the temporal generalization method, the standard *s* is usually presented for several times at the beginning of an experiment. After *s* has been initially presented, the participant receives in each trial a comparison duration $c_i$, as before spaced below and above the standard. After each presentation of a comparison duration, the participant has to judge whether this duration was the same as the standard or different, by responding with $R_{same}$ or $R_{different}$, respectively. Alternatively, the standard and the comparison may be presented in each trial, and the participant is also asked to judge whether the two stimulus durations are equal, $R_{same}$, or not equal, $R_{different}$ (see Birngruber et al. 2014; Dyjas & Ulrich 2014). Table 3.2 contains example data for such an equality judgment task. When the relative frequency of a same response is plotted against comparison duration, an approximately bell-shaped psychometric function emerges. As before, there are various methods available to summarize such data.

### 3.1      *Same-different Model with Constant Standard Deviation*
First, a parametric method has been suggested by Schneider and Komlos (2008). These authors have assumed that subjects base their judgment on the difference $\Delta = C - S$ between the internal representation of the comparison and the standard and respond with $R_{same}$ if $|\Delta + \varepsilon| < \gamma$ and otherwise with $R_{different}$. The parameter $\gamma$ denotes a constant threshold value and $\varepsilon$ the constant error.

TABLE 3.2    *Exemplary outcome of a psychophysical experiment with an equality judgment task. The experiment uses s = 500 ms and nine comparison levels $c_1,...,c_9$ ranging from $c_1$ = 300 ms to $c_9$ = 700 ms. Row $n_i$ shows how often comparison level $c_i$ was repeated during the course of the experiment. The rows $n_{same,i}$ and $n_{different,i}$ give the number of $R_{same}$ and $R_{different}$ responses. The row $f_i$ contains the relative frequencies of the $R_{same}$ responses.*

|                | $c_1$ | $c_2$ | $c_3$ | $c_4$ | $c_5$ | $c_6$ | $c_7$ | $c_8$ | $c_9$ |
|----------------|------|------|------|------|------|------|------|------|------|
|                | 300  | 350  | 400  | 450  | 500  | 550  | 600  | 650  | 700  |
| $n_i$          | 15   | 15   | 14   | 15   | 12   | 15   | 15   | 15   | 14   |
| $n_{same,i}$   | 0    | 5    | 6    | 11   | 8    | 4    | 1    | 0    | 0    |
| $n_{different,i}$ | 15 | 10   | 8    | 4    | 4    | 11   | 14   | 15   | 14   |
| $f_i$          |      | 0.00 | 0.33 | 0.43 | 0.73 | 0.67 | 0.27 | 0.07 | 0.00 | 0.00 |

If one assumes that $\Delta$ follows a normal distribution with mean $\mu_c = c - s + \varepsilon$ and standard deviation $\sigma$, then it can be shown that the probability of a $R_{same}$ response is given by

$$P\left(R_{same} \mid c,s\right) = \Phi\left(\frac{\gamma - (c-s) - \varepsilon}{\sigma}\right) - \Phi\left(\frac{-\gamma - (c-s) - \varepsilon}{\sigma}\right). \tag{11}$$

Again the maximum likelihood method can be used to obtain estimates of $\gamma$, $\varepsilon$, and $\sigma$ from the observed data. The supplementary Matlab script "MLESameDifferent.m" performs this analysis (see book's GitHub repository). Applying this procedure to the data of Table 3.2 yields $\gamma = 62.0$ ms, $Se = 7.1$ ms, $CI = [48.1, 75.9]$, $\varepsilon = -42.9$ ms, $Se = 9.1$ ms, $CI = [-60.6, -25.1]$, and $\sigma = 53.7$ ms, $Se = 7.6$ ms, $CI = [38.8, 68.7]$. Figure 3.5 depicts the resulting psychometric
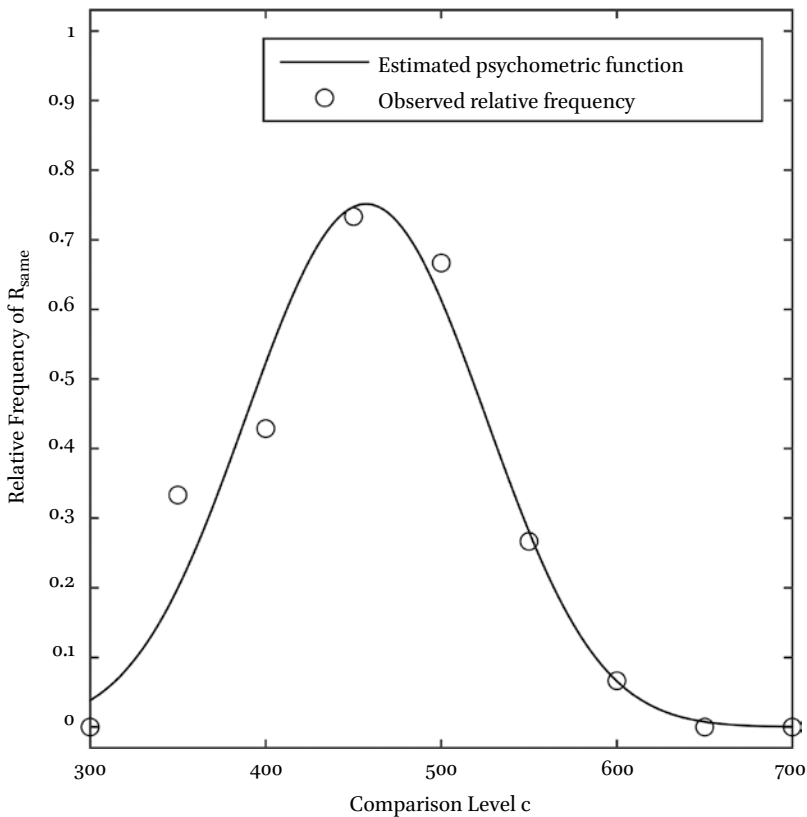


FIGURE 3.5   *Relative frequency of responding with $R_{same}$ (i.e., judging c and s as equally long). The solid line shows the best fitting psychometric function. This model assumes that the standard deviation $\sigma$ of the internal difference $\Delta$ is constant.*

function for these parameters. Note that the estimate of $\sigma$ can be regarded as a sensitivity measure and thus might be reexpressed as $DL = \sigma \cdot z_{0.75}$. Likewise, the $PSE$ can be obtained via $PSE = s + \varepsilon$. For the present example, this yields $PSE = 457.1$ ms, $Se = 9.1$ ms, $CI = [439.4, 474.9]$, and $DL = 36.3$ ms, $Se = 5.1$ ms, $CI = [26.2, 46.3]$.

### 3.2     Same-different Model with Standard Deviation Dependent on Comparison Level

The model underlying Equation 11 implies a symmetrical bell-shaped psychometric function. However, experiments employing the temporal generalization method or the standard procedure of presenting $s$ and $c$ in fixed order in each trial, typically generate asymmetrical psychometric functions with a positive skew (Birngruber et al. 2014; Wearden et al. 1998; Wearden 1992). In order to account for this asymmetrical shape, one may as before (i.e., Pseudo-Gaussian Model) assume that the standard deviation $\sigma$ in the preceding Equation 11 increases with the comparison level $c$, i.e., $\sigma_c = w \cdot c$ (see Birngruber et al. 2014),

$$P\left(R_{same} \mid c,s\right) = \Phi\left(\frac{\gamma - (c-s) - \varepsilon}{w \cdot c}\right) - \Phi\left(\frac{-\gamma - (c-s) - \varepsilon}{w \cdot c}\right). \tag{12}$$

Figure 3.6 displays the estimated function for this model variant when it is applied to the example data in Table 3.2. The parameter estimates, derived by the supplementary Matlab script "MLESameDifferent2.m" (see book's GitHub repository), are $\gamma = 61.5$ ms, $Se = 7.1$ ms, $CI = [47.7, 75.3]$, $\varepsilon = -52.3$ ms, $Se = 9.0$ ms, $CI = [-69.9, -34.6]$, and $w = 0.118$, $Se = 0.016$, $CI = [0.086, 0.150]$. Because the predicted shape of this psychometric function is asymmetrical and influenced by the Weber fraction, it is difficult to properly define a measure of $PSE$. However, similar to the previous definition, one may again compute $PSE = s + \varepsilon$. Discrimination sensitivity is reflected in the parameter $w$, i.e., the Weber fraction. Due to the asymmetry of the underlying psychometric function, this measure should be used to index sensitivity.

### 3.3     Waveform Moment Analysis

The preceding two procedures involved a parametric approach to the analysis of data emerging from equality judgments. The *Waveform Moment Analysis* enables a non-parametric approach (Cacioppo & Dorfman 1987). Let $f_i$ be the observed relative frequency of a $R_{same}$ response associated with comparison level $c_i$. In a first step, these frequencies are converted to a probability distribution $p_i$, $i = 1,...,k$, by the following transformation,
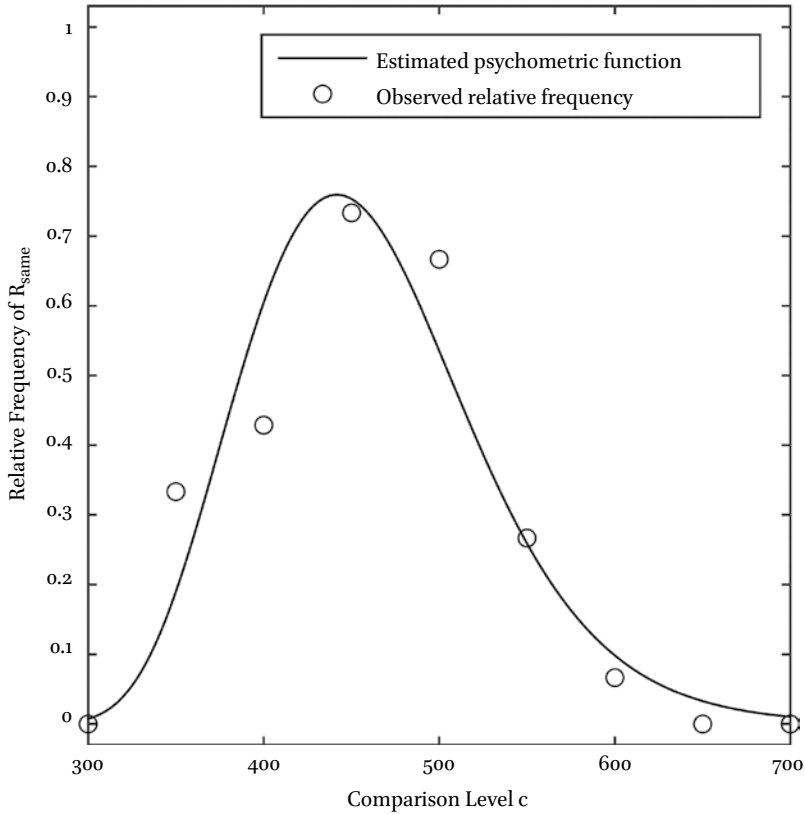
$$p_i = \frac{f_i}{\sum_{i=1}^{k} f_i}. \qquad (13)$$

In a second step, the mean $\mu$ and the standard deviation $\sigma$ are computed for this "probability distribution", that is,

$$\mu = \sum_{i=1}^{k} p_i \cdot c_i \qquad (14)$$

and

$$\sigma = \sqrt{\sum_{i=1}^{k} p_i \cdot (c_i - \mu)^2}. \qquad (15)$$

The parameter $\mu$ assesses the location of the psychometric function on the abscissa and thus can be interpreted as PSE, whereas the parameter $\sigma$ captures the spread of this function and thus reflects discrimination performance with smaller values of $\sigma$ indicating a higher level of discrimination sensitivity. Applications of the waveform moment analysis in temporal discrimination have been reported by Birngruber et al. (2014) and by Dyjas and Ulrich (2014). A Matlab script for performing this analysis ("WaveformMoment.m") is available as supplementary material (see book's GitHub repository). For the data in Table 3.2, one obtains $\mu$ = 456.1 ms, $\sigma$ = 63.8 ms, and thus CE = −43.9 ms. This script also computes the standard error and confidence intervals for these parameters by the bootstrap method. For example, one obtains for $\mu$: Se = 8.7 ms, CI = [438.5, 472.5], for $\sigma$: Se = 5.2 ms, CI = [52.6, 72.8], and for CE: Se = 8.7 ms, CI = [−61.5, −27.5] (see also Figure 3.7).
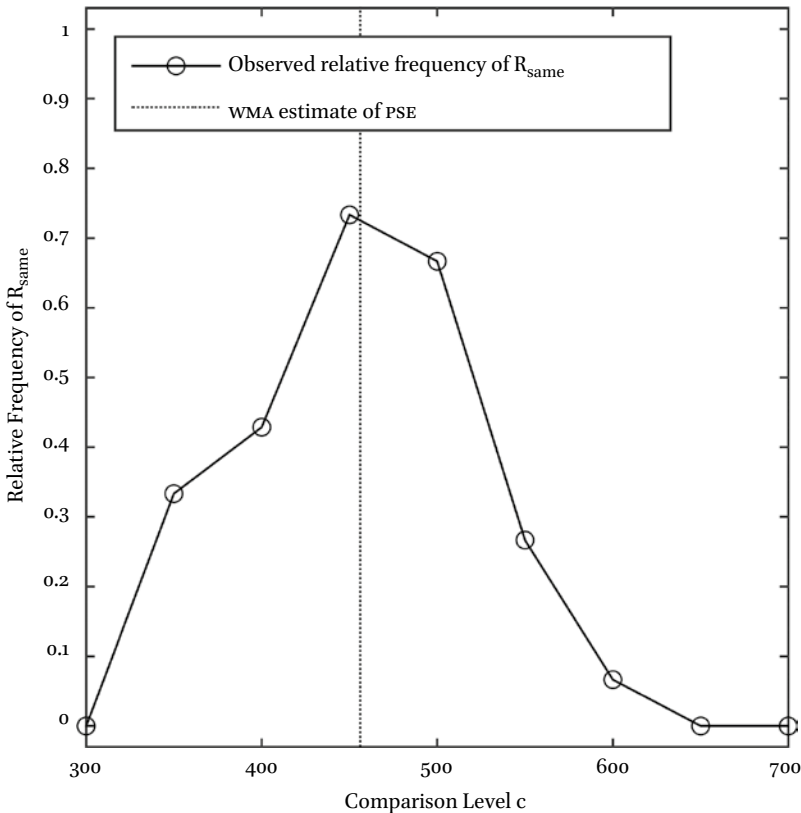


FIGURE 3.7   *Relative frequency of responding with $R_{same}$ (i.e., judging c and s as equally long). The dotted line indicates the PSE estimate derived by means of the Waveform Moment Analysis.*

## 4        Conclusions

In this chapter we have introduced and reviewed several psychophysical paradigms and analytical procedures for determining discrimination sensitivity and perceived duration. For each procedure, we provided Matlab example codes for estimating the respective parameters that beginners to the field of timing research might find helpful (see book's GitHub repository). We have discussed two major approaches differing in the type of judgment employed to assess timing performance, namely, comparative and equality judgments. The two major parameters of interest in this regard are PSE and DL. In experimental work, the absolute magnitude of these parameters is often of subordinate importance. Rather, the major interest lies in assessing differences in these parameters between experimental conditions. For example, an experimenter might be interested in whether or not the size of a visual stimulus affects perceived duration (e.g., Mo & Michalski 1972; Rammsayer & Verner 2014). In this case, the PSE should be estimated for large and for small comparison stimuli, using the same standard in both conditions. When the different size conditions are presented in random order within an experimental block, changes in PSE can be attributed to differences in the size of the comparison stimuli, since other influences on PSE, such as the time-order error, should affect PSE to an identical extent in both conditions. Therefore, a reminder design with fixed order of $s$ and $c$ is usually appropriate whenever an experimenter wants to investigate whether an experimental manipulation affects PSE.

It must be kept in mind, however, that perceived duration still can only be indirectly inferred from changes in the PSE, since the PSE reflects not only changes in perceived duration, but also decisional and response biases, and, therefore, this parameter should be cautiously interpreted in terms of *judged* duration rather than *perceived* duration. The use of a 2AFC task has the additional advantage that one can isolate the effects of secondary experimental manipulations from the time-order error by analyzing the order-conditional psychometric functions. Also, unbiased estimates of DL can be achieved by assessing the slope of the order-conditional functions.

Traditionally, comparative judgments have been used most often to measure both DL and PSE. However, equality judgments may of course also be employed, and might be especially useful to assess the robustness of experimental effects. For example, consider that one is interested in whether an experimental manipulation influences perceived duration. If similar PSE effects can be observed for comparative and equality judgments, this might strengthen the notion that the manipulation affects perceived duration rather than decisional processes (e.g. Birngruber et al. 2014; Dyjas & Ulrich 2014).

Supplementing this chapter, we provided basic Matlab scripts to illustrate the various psychophysical procedures for newcomers to the field of time perception (see book's GitHub repository). It must be mentioned, however, that elaborated psychophysical toolboxes are available for data analysis (see Table 3.3). We also refer the reader to comprehensive manuals on psychophysical methods for information about details of these toolboxes (Kingdom & Prins 2010; Lu & Dosher 2014). For example, the toolbox developed by Wichmann and Hill (2001) also allows the estimation of lapses in designs with comparative judgments. The toolbox *Palamedes* described in Kingdom and Prins (2010) also includes Matlab scripts for adaptive psychophysical procedures. Finally, the Matlab script by Bausenhart et al. (2012) is recommended for fitting psychometric functions conditional on stimulus order in 2AFC tasks.

In this chapter, we focused on psychophysical tools and procedures to obtain and analyze psychometric functions. This is sometimes considered as the classical psychophysical approach. An alternative approach for characterizing discrimination performance is offered by Signal Detection Theory (SDT; Green & Swets, 1966). Interestingly, in the domain of time perception, the psychophysical tools from SDT are much less often used than the classical tools described in this chapter. One major reason why time perception researchers usually prefer the classical tools is that SDT does not provide a parameter like the *PSE* that would allow to estimate judged duration. This is perhaps not surprising since SDT was mainly developed to identify near-threshold stimuli, an issue that does not apply to time perception. Furthermore, we did not address duration scaling methods as temporal reproduction, production, or verbal estimation, which are also often used to investigate duration perception (e.g., Allan 1979; Bindra & Waksberg 1956; see also Chapter 4 of this book). However,

TABLE 3.3 *Overview of advanced toolboxes and functions for psychometric function fitting.*

| Name | Website | References | Notes |
| --- | --- | --- | --- |
| Psychtoolbox | www.psychtoolbox.org | Brainard (1997), Pelli (1997) | See folder "psychometric". Function fitting (e.g., cumulative normal, Weibull, Naka-Rushton). For the cumulative normal and Weibull optimization toolbox is needed. |
| Palamedes | www.palamedestoolbox .org | Kingdom & Prins (2010) | Fits several types of psychometric function also to multiple conditions at the same time, performs model comparison, and contains adaptive procedures. |

TABLE 3.3     *Overview of advanced toolboxes and functions* (cont.)

| Name | Website | References | Notes |
|------|---------|-----------|-------|
| MLE2AFC | http://link.springer.com/ article/10.3758%2Fs13 428-012-0207-z | Bausenhart et al. (2012) | Fits psychometric functions conditional on stimulus order accounting for lapses and response errors. |
| Psignifit | psignifit.sourceforge.net | Fruend, Haenel, & Wichmann (2011) | Performs maximum-likelihood fits, tests the quality of the fit, and provides confidence intervals on the parameters of the fitted functions. |
| UML | http://hearlab.ss.uci.edu/ UML/uml.html | Shen, Dai, & Richards (2015) | Increases efficiency of data collection by estimating the parameters of the psychometric and optimizing stimulus sampling. |
| model free | http://personalpages .manchester.ac.uk/staff/ d.h.foster/software -modelfree/latest/index .html | Zychaluk & Foster (2009) | Non-parametric local linear fitting. |
| PMETRIC | www.psy.otago.ac.nz/ miller/Software.htm | Miller & Ulrich (2004) | WIN EXE that performs probit analysis and Spearman-Kärber method and uses bootstrapping for standard errors of parameter estimates (a Matlab wrapper function is available from the authors). |
| Psychophysica | Available upon personal request to the authors | Watson & Solomon (1997) | Mathematica Notebooks, of which Psychometrica.nb fits and plots psychometric data. |
| quickpsy | dlinares.org/quickpsy .html | Linares & López-Moliner (2016) | R toolbox that fits and plots psychometric functions for multiple conditions. |
| GLMM | mixedpsychophysics .wordpress.com | Moscatelli, Mezzetti, & Lacquaniti (2012) | R toolbox to combine the analysis of the behavior at the level of single subject and population. |

data analyses for these approaches are performed with common statistical measures as mean and standard deviation of the observed data.

In sum, we hope that the present chapter will direct beginners with little or no background in psychophysics to the most important paradigms and

psychophysical methods for assessing discrimination sensitivity and judged duration. The Matlab scripts provided as supplementary material should provide hands-on experience with these methods, although these scripts cannot replace the elaborated toolboxes mentioned above.

### References

Allan, L.G. (1979). The perception of time. *Perception & Psychophysics*, 26, 340–354.

Allan, L.G., & K. Gerhardt (2001). Temporal bisection with trial referents. *Perception & Psychophysics, 63,* 524–440.

Allan, L.G., & J. Gibbon (1991). Human bisection at the geometric mean. Special Issue: Animal timing. *Learning and Motivation*, 22, 39–58.

Ayer, M., H.D. Brunk, G.M. Ewing, W.T. Reid, & E. Silverman (1955). An empirical distribution function for sampling with incomplete information. *Annals of Mathematical Statistics*, 26, 641–647.

Bausenhart, K.M., O. Dyjas, D. Vorberg, & R. Ulrich (2012). Estimating discrimination performance in two-alternative forced choice tasks: Routines for matlab and R. *Behavior Research Methods*, 44, 1157–1174.

Bausenhart, K.M., D. Bratzke, & R. Ulrich (2016). Formation and representation of temporal reference information. *Current Opinion in Behavioral Sciences, 8,* 46–52.

Bindra, D., & H. Waksberg (1956). Methods and terminology in studies of time estimation. *Psychological Bulletin, 53,* 155–159.

Birngruber, T., H. Schröter, & R. Ulrich (2014). Duration perception of visual and auditory oddball stimuli: Does judgment task modulate the temporal oddball effect? *Attention, Perception, & Psychophysics*, 76, 814–828.

Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 443–446.

Bruno, A., I. Ayhan, & A. Johnston (2012). Effects of temporal features and order on the apparent duration of a visual stimulus. *Frontiers in Psychology*, 3, 1–7.

Cacioppo, J.T., & D.D. Dorfman (1987). Waveform moment analysis in psychophysiological research. *Psychological Bulletin*, 102, 421–438.

Dyjas, O., & R. Ulrich (2014). Effects of stimulus order on discrimination processes in comparative and equality judgements: Data and models. *Quarterly Journal of Experimental Psychology*, 67, 1121–1150.

Dyjas, O., K.M. Bausenhart, & R. Ulrich (2012). Trial-by-trial updating of an internal reference in discrimination tasks: Evidence from effects of stimulus order and trial sequence. *Attention, Perception, & Psychophysics*, 74, 1819–1841.

Eisler, H., A.D. Eisler, & Å. Hellström (2008). Psychophysical issues in the study of time perception. In Grondin, S. (Ed.), *Psychology of time* (pp. 75–109). Bingley, UK: Emerald.

Fechner, G.T. (1889). *Elemente der Psychophysik I* (2nd ed.). Breitkopf & Härtel.

Finney, D.J. (1952). *Probit analysis: A statistical treatment of the sigmoid response curve.* Cambridge: Cambridge University Press.

Fründ, I., N.V. Haenel, & F.A. Wichmann (2011). Inference for psychometric functions in the presence of nonstationary behavior. *Journal of Vision*, *11,* 16.

García-Pérez, M.A. (1998). Forced-choice staircases with fixed step sizes asymptotic and small-sample properties. *Vision Research*, *38*, 1861–1881.

García-Pérez, M.A. (2014). Does time ever fly or slow down? The difficult interpretation of psychophysical data on time perception. *Frontiers in Human Neuroscience*, *8,* 415.

García-Pérez, M.A., & R. Alcalá-Quintana (2011). Improving the estimation of psychometric functions in 2AFC discrimination tasks. *Frontiers in Psychology*, *2,* 96.

García-Pérez, M.A., & R. Alcalá-Quintana (2013). Shifts of the psychometric function: Distinguishing bias from perceptual effects. *The Quarterly Journal of Experimental Psychology*, *66*, 319–337.

Gescheider, G.A. (1997). *Psychophysics: The fundamentals* (3rd ed.). Hillsdale, NJ: Erlbaum.

Green, D., & J. Swets (1966). *Signal detection theory and psychophysics* (*rev. ed.*). Los Altos, CA: Peninsula Publishing, reprinted Edition 1998.

Grondin, S. (2001). Discriminating time intervals presented in sequences marked by visual signals. *Perception & Psychophysics*, *63*, 1214–1228.

Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception & Psychophysics*, *72*, 561–582.

Guilford, J. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill Book Company, Inc.

Hegelmaier, F. (1852). Ueber das Gedächtniss für Linear-Anschauungen. *Archiv für physiologische Heilkunde*, *11*, 844–853.

Hellström, Å. (1985). The time-order error and its relatives: Mirrors of cognitive processes in comparing. *Psychological Bulletin*, *97*, 35–61.

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, *49*, 227–229.

Kärber, G. (1931). Beitrag zur kollektiven Behandlung pharmakologischer Reihenversuche. *Archiv für experimentelle Pathologie und Pharmakologie*, *162*, 480–483.

Killeen, P.R., J.G. Fetterman, & L.A. Bizo (1997). Time's causes. In Bradshaw, C.M. & E. Szabadi (Eds.), *Time and behavior: Psychological and neurobiological analyses* (pp. 79–132). Amsterdam: Elsevier.

Kingdom, F.A.A., & N. Prins (2010). *Psychophysics*. Amsterdam: Elsevier.

Köhler, W. (1923). Zur Theorie des Sukzessivvergleichs und der Zeitfehler. *Psychologische Forschung*, *4*, 115–175.

Leek, M.R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics*, *63*, 1279–1292.

Linares, D., & J. López-Moliner (2016). quickpsy: An R Package to Fit Psychometric Functions for Multiple Groups. *The R Journal, 8,* 122–131.

Lord, F.M., M.R. Novick, & A. Birnbaum (1968). *Statistical theories of mental test scores*. Reading, Massachusetts: Addison-Wesley.

Lu, Z.-L., & B. Dosher (2014). *Visual psychophysics*. London, England: The mit Press.

Luce, R., & E. Galanter (1963). Discrimination. In Luce, R.D., R.R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 191–243). New York: John Wiley & Sons.

Matthews, W.J., & W.H. Meck (2016). Temporal cognition: Connecting subjective time to perception, attention, and memory. *Psychological Bulletin*, *142*, 865–907.

Miller, J., & R. Ulrich (2001). On the analysis of psychometric functions: The Spearman-Kärber method. *Perception & Psychophysics*, *63*, 1399–1420.

Miller, J., & R. Ulrich (2004). A computer program for Spearman-Kärber and probit analysis of psychometric function data. *Behavior Research Methods, Instruments, & Computers*, *36*, 11–16.

Mo, S.S., & V.A. Michalski (1972). Judgment of temporal duration of area as a function of stimulus configuration. *Psychonomic Science*, *27*, 97–98.

Moscatelli, A., M. Mezzetti, & F. Lacquaniti (2012). Modeling psychophysical data at the population-level: The generalized linear mixed model. *Journal of Vision*, *12,* 26.

Nachmias, J. (2006). The role of virtual standards in visual discrimination. *Vision Research*, *46*, 2456–2464.

Pelli, D. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Rammsayer, T. (2010). Differences in duration discrimination of filled and empty auditory intervals as a function of base duration. *Attention, Perception, & Psychophysics*, *72*, 1591–1600.

Rammsayer, T., & R. Ulrich (2001). Counting models of temporal discrimination. *Psychonomic Bulletin & Review*, *8*, 270–277.

Rammsayer, T., & R. Ulrich (2012). The greater temporal acuity in the reminder task than in the 2AFC task is independent of standard duration and sensory modality. *Canadian Journal of Experimental Psychology*, *66*, 26–31.

Rammsayer, T., & M. Verner (2014). The effect of nontemporal stimulus size on perceived duration as assessed by the method of reproduction. *Journal of Vision*, *14*, 1–10.

Renz, T., & A. Wolf (1856). Versuche über die Unterscheidung differenter Schallstärken. *Archiv für physiologische Heilkunde*, *15*, 185–193.

Schneider, K.A., & M. Komlos (2008). Attention biases decisions but does not alter appearance. *Journal of Vision*, *8*, 1–10.

Shen, Y., W. Dai, & V.M. Richards (2015). A MATLAB toolbox for the efficient estimation of the psychometric function using the updated maximum-likelihood adaptive procedure. *Behavior Research Methods*, *47*, 13–26.

Spearman, C. (1908). The method of "right and wrong cases" ("constant stimuli") without Gauss's formulæ. *Britisch Journal of Psychology*, *2*, 227–242.

Sternberg, S., R.L. Knoll, & P. Zukofsky (1982). Timing by skilled musicians. In Deutsch, D. (Ed.), *The psychology of music* (pp. 181–239). New York: Academic Press.

Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research*, *35*, 2503–2522.

Treutwein, B., & H. Strasburger (1999). Fitting the psychometric function. *Perception & Psychophysics*, *61*, 87–106.

Ulrich, R. (1987). Threshold models of temporal-order judgments evaluated by a ternary response task. *Perception & Psychophysics*, *42*, 224–239.

Ulrich, R. (2010). DLs in reminder and 2AFC tasks: Data and models. *Attention, Perception, & Psychophysics*, *72*, 1179–1198.

Ulrich, R., & J. Miller (2004). Threshold estimation in two-alternative forced-choice (2AFC) tasks: The Spearman-Kärber method. *Perception & Psychophysics*, *66*, 517–533.

Ulrich, R., & D. Vorberg (2009). Estimating the difference limen in 2AFC tasks: Pitfalls and improved estimators. *Attention, Perception, & Psychophysics*, *71*, 1219–1227.

Watson, A.B., & J.A. Solomon (1997). Psychophysica: Mathematica notebooks for psychophysical experiments. *Spatial Vision*, *10*, 447–466.

Wearden, J.H. (1992). Temporal generalization in humans. *Journal of Experimental Psychology: Animal Behavior Processes*, *18*, 134–144.

Wearden, J.H., P. Rogers, & R. Thomas (1997). Temporal bisection in humans with longer stimulus durations. *Quarterly Journal of Experimental Psychology*, *50B*, 79–94.

Wearden, J.H., H. Edwards, M. Fakhri, & A. Percival (1998). Why "sounds are judged longer than lights": Application of a model of the internal clock in humans. *The Quarterly Journal of Experimental Psychology*, *51B*, 97–120.

Wichmann, F.A., & N.J. Hill (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*, 1293–1313.

Woodworth, R.S., & H. Schlosberg (1954). *Experimental psychology* (*3rd ed.*). London: Methuen.

Żychaluk, K., & D.H. Foster (2009). Model-free estimation of the psychometric function. *Attention, Perception, & Psychophysics*, *71,* 1414–1425.