[All URLs appearing in this chapter have been verified. CE]

<rrh>Psychophysics: Concepts, Methods, and Frontiers

<vrh>*Hexley, Morimoto, and Spitschan*

<ack>https://doi.org/10.1037/0000318-024
APA Handbook of Research Methods in Psychology, Second Edition: Vol. 1. Foundations, Planning, Measures, and Psychometrics, H. Cooper (Editor-in-Chief)
Copyright © 2023 by the American Psychological Association. All rights reserved.

<cn>24

<ct>**Psychophysics: Concepts, Methods, and Frontiers**

<au>Allie C. Hexley, Takuma Morimoto, and Manuel Spitschan

<ack>A.C.H. is supported by funding received under the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement N° 765911 (RealVision). T.M. is supported by the Wellcome Trust (Sir Henry Wellcome Fellowship to T. M.; Wellcome Trust 218657/Z/19/Z) and Pembroke College, University of Oxford (Junior Research Fellowship to T. M.). M.S. is supported by the Wellcome Trust (Sir Henry Wellcome Fellowship to M.S.; Wellcome Trust 204686/Z/16/Z).

<h1>**What Is Psychophysics?**

<bt>Fundamental to the human experience is sensing and perceiving properties of the environment. In natural language, *sensing* generally refers to the processing of physical stimuli and patterns in our environment and the extraction of relevant information. As psychologists, we want to understand these processes systematically and scientifically. How can we quantify the quality or magnitude of a percept, and how can we meaningfully compare percepts between

different individuals or from different environments? While we can easily measure physical phenomena, such as the weight of an object in grams using the appropriate instruments (e.g., a scale), the measurement of perceptual phenomena is nontrivial. The set of methods in psychological science that concern the measurement of sensory and perceptual phenomena through behaviour is called *psychophysics*. The goal of psychophysics is to quantitively characterize the relationship between the physical properties of a stimulus—such as the radiance, spatial frequency content, or motion velocity for a visual stimulus, or the amplitude, pitch, or spatial location for an auditory stimulus—and our associated percept.

Psychophysical methods have a long history in psychological science. A major milestone is the formalization of psychophysics of Gustav Theodor Fechner (1801–1887) in the 1860s, though the study of perception, of course, does not start in the 19th century but has a much longer history (Wade, 1998). Since then, psychophysical methods have contributed to discovering numerous fundamental rules that appear to govern our sensory and perceptual processes. Psychophysical experiments can take many forms, yet a common denominator is usually the precise and parametric control of stimuli along one or multiple physical dimensions.

As such, experiments are intimately linked to the technology available for stimulus control. In particular, the arrival of computers and displays in the 1960s and 1970s stimulated psychophysical research. For example, modern computer graphics methods have dramatically increased the ability to control visual stimuli and make them more naturalistic. Furthermore, while the traditional paradigm for psychophysical experiments is a laboratory setting, online experiments have facilitated the collection of psychophysical data from hundreds or thousands of participants in more recent years. Psychophysics is very much alive and indeed provides valuable insights into human sensory and perceptual processing.

The psychophysical literature is immense. It is hopeful and naïve at best and impossible at worst to attempt to summarize it or give a coherent definition for it within the scope of a single book chapter. Therefore, we do not engage with the premise of psychophysics in this chapter: Sensation and perception are indeed objectively measurable, which has been extensively detailed elsewhere (Box 24.4, Further Reading, for pointers at the end of the chapter). We provide a pragmatic approach to psychophysics. First, we discuss why psychophysics is useful and present examples from the published literature of psychophysics "in action." We will then examine and discuss the ingredients of a psychophysical experiment: the measure, the stimulus, the observer, the task, the method, and the analysis. Finally, we discuss the current frontiers of psychophysical experimentation.

At the outset, we note that our approach to psychophysics is fundamentally one focused around visual psychophysics (as our backgrounds are all rooted in vision science). We, therefore, primarily discuss examples from vision and visual perception. The key ideas we discuss, however, are applicable across all perceptual mechanisms, including, for example, audition, somatosensation, taste perception, and multisensory perception.

<h1>**Why Is Psychophysics Useful?**

<bt>At the core of psychophysics is the measurement and characterization of human sensory and perceptual responses to external stimuli. Psychophysical methods are useful for psychological science for a variety of reasons: (a) to quantify and put bounds on human sensation and perception, (b) as a tool to predict underlying sensory and perceptual mechanisms, (c) to inform neural underpinnings, and (d) as a tool to guide ergonomic and “human-centric” design.

<h2>**Psychophysics as a Tool to Quantify and Put Bounds on Human Sensation and**

Perception <bt>Psychophysical measurements help us understand the properties and limitations of human sensation and perception. One of the operative units of analysis of psychophysics is the concept of the threshold: the minimum level of a physical stimulus at which a specific perceptual performance level occurs. Threshold measurements allow us to quantify the bounds of our perception. In many cases, psychophysics allows us to characterize when physically distinguishable stimuli are perceptually indistinguishable. For example, color discrimination thresholds tell us the amount of physical difference that needs to exist between two different colors for them to be reliably distinguished as separate colours. Importantly, psychophysical methods are also at the basis of many vision and hearing examinations in a clinical context (Johnson, 2013).

<h2>Psychophysics as a Tool to Predict Underlying Sensory and Perceptual Mechanisms <bt>In addition to describing and characterizing performance, psychophysical measurements can also uncover underlying sensory and perceptual mechanisms or classes of mechanisms when designed carefully. One specific tool that allows for this is adaptation, also termed the *psychophysicist's electrode* (Frisby, 1980). The logic underlying adaptation is as follows. An observer is repeatedly exposed to a specific stimulus, the adapting stimulus. If this exposure modifies performance on a given task (but not another), this means that the adapting stimulus has affected a mechanism underlying this specific task. In other words, "if it adapts, it's there" (Webster, 2015). A particular example in which adaptation has been used to uncover specific mechanisms is color adaptation: Exposure to an adapting sinusoidal stimulus modulating along one dimension of color space modifies performance along that dimension but not along others (Krauskopf et al., 1982).

<h2>**Psychophysics Helps Inform Neurophysiological Underpinnings** <bt>While

psychophysics fundamentally provides a behavioural output, psychophysical data can inform the underlying neurophysiological processes that ultimately govern behavior and performance (even though these are important in their own right). In many ways, neurophysiological data describing a specific neural response to a physical stimulus are more compelling when supported by psychophysical data.

<h2>**Psychophysics as a Tool to Guide Ergonomic and “Human-Centric” Design**

<bt>In addition to these basic uses of psychophysical measurement techniques, which ultimately help develop our mechanistic understanding of how the brain processes information from the external world, psychophysics also generates knowledge that is immediately useful in the world, and there are many examples of this in our day-to-day life. For example, computer monitors have only three independent primary colors—red (R), green (G), and blue (B)—because human color vision is three-dimensional, owing to the trichromatic retina comprising three classes of cones: long (L), medium (M), and short (S). Therefore, additional primaries would carry no further color information for the observer to exploit. Knowledge of the trichromatic nature of human colour vision, therefore, directly informs display colour reproduction design. When you look for it, you notice that many aspects of design in the human environment are ultimately guided by psychophysical data, such as the choice of specific warning signals and colors and the development of JPEG and MP3 standards for image and sound compression, respectively.

<h1>**What Are the Ingredients of a Psychophysics Experiment?**

<bt>The preceding section has hopefully illustrated the benefits of psychophysics and convinced the reader that psychophysics is still very much a valuable tool in psychological science today. This section provides an outline of the key considerations an experimenter has to make when designing a psychophysics experiment. Following a modified framework from

Kingdom and Prins (2016), we break this down into the “six ingredients” of a psychophysics experiment: the measure, stimulus, the observer, the task, the method, and the analysis.

<h2>***Measure***

<bt>The first thing to consider when designing a psychophysical experiment is what it is that we are trying to measure with our experiment. Any model parameters estimated from psychophysical data can be taken as a psychophysical measure. The exact measure of interest is inherently tied to the psychophysical experimental design. We do not unpack all possible measures one could draw from psychophysical data but rather highlight a few common ones and how one may determine an appropriate measure for their data. One clear dichotomy to emphasize when considering which measure to use is whether the measure is of a *threshold* or some *suprathreshold* metric.

A pervasive psychophysical measure, the threshold quantifies the limits of our perceptual ability—e.g., the threshold for detecting a stimulus embedded in noise or the absolute threshold of vision. Thresholds are points along a specific physical dimension for which particular levels of psychophysical performance are defined. Sensitivity, defined as the reciprocal of the threshold, is another commonly used measure. The *sensitivity index, d'* (dee-prime), is the standardized psychophysical measure for sensitivity (Box 24.2, Signal Detection Theory).

Thresholds can be further subdivided into *detection thresholds* and *discrimination thresholds.* Detection thresholds refer to the minimum level of a stimulus required for sensing the presence of the stimulus, such as detecting the absolute threshold of vision. Discrimination thresholds refer to the minimum difference between two stimuli required for the observer to be able to distinguish the two stimuli as different from each other, such as the minimum color difference between two stimuli required for the two to be perceived as different colors. The *just-*

noticeable difference (JND) is a measure of the difference in physical stimulus intensity required for a difference to be perceived, while the *point of subjective equality* (PSE) is a measure of the physical stimulus range over which the two stimuli are perceived to be identical to each other.

Suprathreshold experiments are primarily focused on measuring perceptual performance above threshold level, for example, measuring the perceived loudness over a wide range of sound amplitudes, rather than looking for the amplitudes at which a sound can be detected or discriminated from another. Suprathreshold measures tend to be a *sensory scale*, relating the physical stimulus magnitude to the perceived magnitude. This relationship is typically characterized by Stevens' power law (Box 24.1, Classical Psychophysical Laws), and the measure would be the constants that describe the mathematical relationship between the physical and perceived magnitude specific to the given modality.

<h2>***Stimulus***

<bt>A fundamental requirement for a psychophysical experiment is to be able to control something in the environment in a given modality and present it to an observer in some systematic fashion. This "something" is called the *stimulus* (pl. *stimuli*). Usually, stimuli are defined along some physical dimension that can be characterized using physical measurements such that they have a "ground truth" value, and stimuli are parametrized along this specific physical dimension. For example, we can present and modulate the radiance or the spectral content (i.e., the wavelength composition) of light using well-defined physical measurements and use these as stimuli to understand our perception of brightness and color. A key principle of stimulus design is *stimulus reduction* (Koenderink, 1999), that is, reducing the rather complex visual world into parametrizable components.

In practice, this control and parametrization of stimuli cannot be achieved using everyday

objects but requires dedicated apparatus for it. Prior to the advent of personal computers, which allow for the relatively easy generation of perceptual stimuli with a good amount of user control, psychophysical experiments were generally performed with custom-made equipment tailored to fit the needs of the stimulus and investigation in question.

Generally, psychophysical experiments with their high demands on stimulus control are performed in laboratory suites in which other aspects that might interfere with the presentation of stimuli—such as room illumination or glare from windows that can interfere with the presentation of visual stimuli on a display—can be minimized. These conditions lend themselves to restrictive viewing conditions (Koenderink, 1999) to ensure that truly only one parameter, the one under investigation, of the physical world is changed between trials and across an experiment while minimizing possible behavioural confounders such as head movement (controlled by a chin rest and bite bar) or eye movement (controlled by instructions and conjoint eye movement registration).

In visual psychophysics, we distinguish between *Newtonian* (free) viewing conditions and *Maxwellian* viewing conditions. In Maxwellian viewing conditions, stimuli are imaged in high-contrast directly onto the pupil of the observer, requiring special focusing optics. The majority of contemporary psychophysicists use Newtonian (free) viewing conditions with computer monitors to display visual stimuli and manipulate their properties.

Psychophysical data can be collected in museums (Martin, 2014), in extreme environments (Willmann et al., 2010), and during Guinness world record attempts (Todor, 1975). Additionally, online experiments have recently opened new means for psychophysical data collection (Sauter et al., 2020), they but present problems for stimulus calibration and precise control. When presented online, visual stimuli are not calibrated in size, luminance, color, and

timing and, therefore, may not yield high-quality results. Workarounds have been proposed to address some aspects of this, for example, using a common object such as a credit card for calibrating size and distance (Li et al., 2020) or using perceptual calibration techniques for display linearization (Xiao et al., 2011). Another promising way to deliver stimuli (including in remote experiments) may be using consumer-grade virtual reality (VR) sets (Scarfe & Glennerster, 2019).

A key task of the psychophysicist is to ensure that stimuli are calibrated. Calibration refers to the practice of ensuring that the physical dimensions one hopes to manipulate and control in a psychophysical experiment are indeed under full control. This is to ensure that aspects of psychophysical performance can be uniquely ascribed to the psychophysical observer and to rule out any confounding artifacts.

For computer-controlled monitors, this involves the measurement of the light emitted from the monitors using a spectroradiometer or colorimeter (Brainard et al., 2002) and using these physical measurements to inform which parameters need to be used to produce a specific stimulus of known characteristics. Spatial and temporal characterization is another necessary step of monitory calibration. The exact nature of calibration depends on the modality. For instance, in visual psychophysics, calibration is focused on characterizing the spectral, spatial, and temporal content of the light and often performed using specialized physical equipment to measure the physical properties of the light. In auditory psychophysics, however, calibration is often done perceptually: using experienced observers to match the perceived loudness of two sounds that differ in a different physical attribute of interest, such as pitch. Developing confidence in the physical characteristics of the stimulus, whether through physical or perceptual calibration, is a key and nontrivial aspect of psychophysical stimulus design and development.

Just as with the measure, we often distinguish between stimuli at threshold and suprathreshold stimuli. Threshold stimuli are those that are supposed to be difficult to detect and distinguish. Suprathreshold stimuli are those that are above a given psychophysical threshold.

<h2>***Observer***

<bt>If the goal of psychophysics is to relate external phenomenon to perceptual experiences, then in addition to our calibrated, well-defined physical stimulus, we clearly also need an observer to sense the stimulus and report their perceptual experience. The observer is, thus, the participant in the psychophysical experiment, whose performance and perception are studied (in contrast with the "experimenter" who runs the experiment). In this chapter, we focus on psychophysics when the observer is human, but, in principle, all organisms with overt behaviour—or indeed, computational algorithms—can be psychophysical observers.

Just as one needs to consider the characteristics of their stimuli, one must also consider the characteristics of their observer. For instance, in visual psychophysics, the experimenter needs to know if their observers have normal visual acuity and typically whether or not they have a color vision deficiency. Comparing results in color matching tasks between trichromats and dichromats without knowing the observer's color deficiency status would lead to confusing results. Additionally, the experimenter can often choose to manipulate the state of the observer for their experimental advantage.

A commonly used example of such observer state manipulation is adaptation, and a psychophysical experiment often begins with an adaptation procedure. The precise nature of this procedure depends on the research question studied. In studies of human vision, a common procedure is dark adaptation. During dark adaptation, an observer is in the dark, with no light available. The idea behind dark adaptation is to sensitize the rods, the retinal photoreceptors

sensitive to dim light (in contrast to the cones, which are active in daylight) and, thereby, remove any lingering “photic history” effects due to the slow recovery of rods. Adaptation to the background, that is, presentation of a neutral stimulus in the absence of a task-related stimulus, is often performed to ensure that stimuli are indeed well-defined in contrast.

Generally, a distinction is made between naïve and trained observers. The term *naïve* here is used in its technical sense, as in not knowing the purpose of an experiment or the hypotheses it tests. In practice, every psychophysical experimenter is typically an observer in their own experiment at least once, to test it and make sure it runs smoothly. In psychophysics, self-testing and self-experimentation has a long history, with many authors describing themselves as “experienced psychophysical observers” and reporting their own data in experiments. This is not least because many psychophysical experiments are very demanding and time-consuming (and frankly, not often the most exciting way to spend a few hours of your day).

In threshold psychophysics, experienced observers may be preferred to naïve observers, as threshold psychophysics aim to test the limits of sensation. There is typically, as with any psychological experiment, an initial period in which performance on psychophysics tasks improves before reaching peak performance, that is, the threshold level. Experienced observers may require less time to reach such a peak performance than naïve observers. Participants who have never participated in a psychophysical experiment (or quite often, every participant) may benefit from a set of practice trials at the start of the experiment, which simulates the task without leading to archival data.

Unlike most psychological experiments, psychophysical experiments have typically relied on collecting large volumes of data from a very small number of observers. The use of

study designs with small samples has a long history and a prominent place in current practice (Smith & Little, 2018). This is often supported by the desire to use experienced observers who require minimal instruction and no further external motivation, as discussed above. A different paradigm is the use of a large number of observers with fewer trials collected per participant. While these are often impractical or hard-to-resource in the laboratory, they provide valuable insights into individual differences (Mollon et al., 2017).

The use of online experiments and online recruitment platforms such as Prolific or Pavlovia has facilitated accessing larger numbers of observers with fewer resources (time spent with and recruiting participants, participant payments for in-laboratory visits). It is, however, imperative to consider the data ethics surrounding the use of these platforms, which may sometimes be labor markets rather than participant pools (Fort et al., 2011; Moss et al., 2020).

<h2>***Task***

<bt>Just as the visual world is reduced into parametrizable and controllable dimensions, visual behavior is also reduced in a psychophysical experiment. Rather than allowing “natural” and uncontrollable behavioral responses to visual stimuli, in a psychophysical experiment, behavior is subject to a specific task that the participant performs. Broadly speaking, a task is any type of behavior that the observer is asked to perform in response to a certain stimulus presentation protocol. There are many different psychophysical tasks, and which to choose is often dictated by the research question asked. As with stimuli and measure, the appropriate task also often depends on whether the experiment is at threshold or suprathreshold levels.

A classic task is a *yes–no* task in which the stimulus is either presented or not on a given trial, and the participant is asked to indicate (e.g., using a button-press) whether or not they sensed the stimulus in that trial. Yes–no tasks are commonly used in high-threshold theory

psychological experiments but are susceptible to response bias in low-threshold theory psychophysical experiments (Box 24.2. Signal Detection Theory). To control for response bias in yes–no tasks, experimenters often collect confidence ratings alongside the observer's response or often avoid such tasks altogether in preference for a bias-free task—an *alternative forced-choice* (AFC*) task.*

In AFC tasks, the observer is presented with multiple spatial locations simultaneously or multiple temporal intervals successively, one of locations or intervals will always contain the stimulus while the others do not, and the observer must respond in which location or interval the stimulus was present. The most common AFC tasks are *two-alternative forced-choice tasks* (2AFC) where two intervals are presented in either distinct spatial locations or successively in time during a single trial, and only one of these intervals contains the stimulus. As long as the stimulus presentation interval is randomized, AFC tasks are inherently free from response bias.

AFC tasks are commonly used for threshold measures. For instance, one might present a Gabor stimulus in one interval and not the other when measuring contrast sensitivity thresholds. AFC tasks are also a popular choice for suprathreshold measurements, where they are often referred to as *pairwise comparison tasks.* In a classic pairwise comparison task, observers are shown stimuli in pairs and asked to rate which stimulus is say brighter, more translucent, or louder. Such judgements are called *ordinal judgements* as the observer's task is to respond with which order these stimuli fall in on the perceptual scale. Ordinal judgements may also be made in *rank order* tasks, where the observer is shown more than two stimuli at once and asked to rank them along the perceptual dimension.

Recently, AFC *difference scaling* tasks are being used instead of classical pairwise comparisons, where the observer is presented with two pairs of stimuli and asked which of the

two pairs has the biggest perceptual difference (Maloney & Yang, 2003). Such *interval judgements* allow the experimenter to extract not only the *ordinal perceptual scale* but a full perceptual *difference scale.*

Suprathreshold stimuli can also be characterized using *category scaling*, one of the oldest psychophysical tasks. Observers are asked to assign stimuli, of a known physical stimulus magnitude, to one of a discrete and finite number of categories. For example, one may be asked to assign chili peppers (with known varying degrees of capsaicin) into one of five categories, with 1 being the least spicy, and 5 the spiciest. Category scaling is limited in that only a set number of categories are used to quantify the perceptual experience, which may not accurately capture the full range of the perceptual experience. For instance, in the chili pepper example, there may be a wide range of spice levels within the least spicy group, which may be perceptually distinct: information that is lost through category scaling.

Magnitude estimation tasks can be used instead of category scaling, where the observer is asked to assign any number to quantify the stimulus, thus opening up the range of possible numerical values the observer can assign to a particular physical stimulus. For example, the experimenter may give the participant a slight electric shock and ask them to put a number to the pain induced by the shock. The experimenter often provides an arbitrary reference value, for example, deliver a shock of 10V and tell the participant that they have been given a shock of level 5. When delivering a shock of 20V, the participant will then freely assign a number to that shock level, in reference to the level 5 shock they received before. Different participants will give different numbers for different levels. For example, one participant may rate the 20V shock a 5.5, while another may rate it a 7, and a third a 20. Whilst the numerical scales assigned differs between participants, scale values are usually repeatable and robust within a participant (there

are some stimuli, however, where participants always struggle to ascribe reliably meaningful magnitude values).

<h2>***Method***

<bt>While we refer to the *task* as what happens on a single experimental trial, we call the *method* the way that the entire psychophysical experiment is organized and how stimulus levels are chosen and how this choice might depend on previous decisions made by the participant. We distinguish between “adaptive” and “nonadaptive” or “fixed” procedures. Broadly speaking, in adaptive methods, namely the method of adjustment or staircases, stimulus levels on a given trial are informed by previous decisions. In nonadaptive procedures, namely the method of constant stimuli or the method of limits, a fixed set of stimuli or stimulus comparisons are shown to the participant.

≤h3≥*Method of Constant Stimuli* <bt>In the method of constant stimuli, stimuli at different magnitudes are presented in randomized order, and the participant is asked to perform the task, such as a 2AFC task detecting a sound in one of two intervals presented successively on each presentation interval. The magnitude of the stimulus should range between the level at which the participant can never detect the stimulus and the level at which participants can easily detect (i.e., a 100% detection rate), with stimuli ideally being spread to induce even detection rates between these extremes. This method is a classic and reliable psychophysical method. One challenge with the method of constant stimuli is knowing a priori the appropriate range of stimulus magnitudes to set. Experiments run using the method of constant stimuli also tend to be time-consuming.

<h3>*Method of Limits* <bt>In the method of limits, the stimulus magnitude is increased or decreased on subsequent trials until a threshold level is reached. Two stimulus presentation

series are typically used: an ascending series in which the starting stimulus magnitude is way below threshold and the stimulus magnitude is increased over subsequent trials until threshold is reached, and a descending series where the starting stimulus magnitude is way above threshold and decreased over subsequent presentations until threshold level is reached. Typically, participants complete both series multiple times, and the threshold is taken as the average of the threshold estimate obtained in each series.

The method of limits tends to be quite quick for data collection, but there are two types of error associated with this method. Firstly, because participants are exposed to similar stimuli many times, they could get used to stimuli (error of habituation). Second, because stimuli intensity gradually changes, some participants might predict when the stimuli will become visible even when they do not see the stimuli (response bias). Careful instruction partially helps suppression of these two sources of errors, though the method of limits is often regarded as being a less accurate measure of thresholds than other methods.

<h3>Method of Adjustment <bt>The method of adjustment gives participants control of the magnitude of stimulus. For example, when measuring absolute thresholds, participants are asked to change the stimulus intensity until they can just detect the stimulus. Trials end when a participant makes a satisfactory match. The initial point of stimulus magnitude is randomized to minimize the error of habituation.

This conceptually simple method is efficient and has the advantage that it requires participants to engage actively in the experiment. It can also often be quicker than requiring the participant to view multiple stimulus magnitudes that are above or below threshold (though this depends on the individual participant; for the more hesitant observers, a time limit may be introduced to encourage them to not spend too long making a match).

<h3>*Staircase method* <bt>The staircase method is an adaptive variation of the method of limits. Over subsequent trials, the stimulus magnitude is updated according to participants' response in a previous trial. In the simple-up down staircase, the stimulus magnitude in the current trial increases by a predefined step size if the participant makes a mistake in the previous trial, and the magnitude decreases when the participant makes a correct response. This procedure continues until the predefined number of reversals (i.e., transition from incremental stream to decremental stream or vice versa) are recorded. Threshold estimates are obtained by averaging the stimulus magnitude across a certain number of reversals (for instance, if the staircase is allowed to reverse 12 times, the last four reversals may be averaged to find the threshold).

Adaptive versions of the staircase method have offered improvements on the classic staircase design. In adaptive staircases, the step size can change over trials, allowing the staircase to settle (or converge) quicker, improving efficiency (Kontsevich & Tyler, 1999; Watson & Pelli, 1983). The next stimulus magnitude presentation is estimated from adaptive staircases based on the history of participants' responses during the experiment and aims to present a stimulus magnitude around the estimated threshold that supports the efficient convergence of the staircase. The idea behind this approach is to make use of the history of observer responses containing rich amount of information about participants' sensitivity.

<h2>***Analysis***

<bt>The final stage in a psychophysical experiment is the analysis. To extract the relevant measure of observer perceptual performance to a specific physical stimulus on a given task and experimental method, one needs to apply the appropriate analysis. The choice of stimulus, task, and method is intimately linked to the choice of analytic procedure applied to psychophysical data. Importantly, different psychophysical experimental designs (to extract

different measures) require different analyses. Kingdom and Prins (2016) review the major classes of psychophysical experiments and provide a toolbox to analyze psychophysical data in MATLAB, and Knoblauch and Maloney (2012) provide a textbook and toolbox on how to analyze psychophysical data in R. Here, we only touch upon a few basic, theoretical analysis concepts. Just as we distinguish between threshold and suprathreshold measures and stimuli, so do analysis techniques distinguish between extracting threshold and suprathreshold measures.

In threshold experiments, the fundamental representation of psychophysical data in detection and discrimination is the *psychometric function*. The psychometric function relates the underlying stimulus dimension that is changed in a psychophysical experiment—for example, brightness of a light or the contrast of stimulus—to the performance of the psychophysical observer (typically plotted as the proportion of correct detections or discriminations). Sensory mechanisms are inherently noisy: meaning that the same level of the physical stimulus may sometimes induce a perceptual response and sometimes may not. Due to this inherent internal noise, the psychometric function has a sigmoid shape, rather than a simple step-function shape that it would take if the observer's sensory system were not subject to internal noise (see Box 24.2. Signal Detection Theory). We do not detail methods for fitting a psychometric function to psychophysical data here but refer the reader to the MATLAB and R toolboxes developed and published by Prins and Kingdom (2018) and Knoblauch and Maloney (2012), respectively, for tools with which to do this.

Psychometric functions are the fundamental analysis method in threshold experiments, yet the sigmoid shape of the psychometric function poses an obvious problem for measuring thresholds—what point on the psychometric function does one take as the threshold? By convention, thresholds are taken as the 50% correct level, though this is an arbitrary choice made

by the experimenter.

Further, when dealing with threshold experiment analysis, one must consider how the observer sets their internal threshold for detection and discrimination. In the Task section, we described the two types of tasks for threshold experiments: yes–no tasks and alternative-forced choice tasks. In a yes–no task, the observer must (typically unconsciously) decide a certain level of internal sensory system response to ascribe to detection or discrimination, such that anytime the observer experiences an internal response greater than that level they respond that they can detect or discriminate the stimulus, and any response below that threshold they respond seeing nothing. Different observers will have different internal thresholds for responding yes or no, known as *response bias*. Thus, if a yes–no task is used, response bias must be separated from sensitivity during analysis. Signal detection theory provides a framework for modelling such inherently noisy sensory mechanisms and untangling thresholds from response bias (see Box 24.2. Signal Detection Theory). By design, alternative-forced choice tasks are criterion bias–free, and, thus, response bias does not need to be untangled during analysis. There may still be, for example, attentional biases, but these are compensated for by randomizing the stimulus presentation interval.

When adaptive methods are applied, psychometric functions do not necessarily need to be used for threshold analysis. For example, in experiments that use staircases, one may simply take the threshold to be the average of the last few reversal points. In method of adjustment experiments, one typically just takes the threshold to be the point set by the observer: the observer typically completes the experiment at least twice, once starting above threshold and once below such that the threshold is taken as the average of the two settings made when approaching the threshold from opposing directions.

For suprathreshold measures, the appropriate analysis method is usually determined based on a combination of the task and the stimulus. Again, different tasks require different analysis methods, and we do not detail individual analysis techniques here but outline the basic principles. Magnitude estimation tasks and category scaling are typically analyzed by fitting a *scaling function,* typically in the form of Steven's power law, to the relationship between the physical stimulus value and the assigned perceptual magnitude value.

In suprathreshold AFC pairwise comparison tasks, the experimenter uses *difference scaling* to extract the difference measure of interest. A multitude of difference scaling approaches exist, and the appropriate difference scaling analysis to use depends on both the physical properties of the stimulus and the type of pairwise comparison task. Classically, for ordinal judgements, Thurstonian scaling (using Thurstone's law of comparative judgement) was applied for difference scaling analysis. Thurstonian scaling is useful in that it can be used for complex stimuli whose physical characteristics are multidimensional and difficult to parameterize, such as judgements on the attractiveness of a painting. Thurstonian scaling can also be used to analyze interval (difference) judgements, though stochastic methods developed specifically for interval judgements—namely *maximum likelihood difference scaling* (MLDS) (Maloney & Yang, 2003), extending similar but nonstochastic methods first developed by Schneider et al. (1974) and Schneider (1980)—are better suited for deriving suprathreshold difference scales.

Further, in suprathreshold tasks where two or more physical dimensions contribute to a single perceptual dimension, for example, scattering and absorption properties of a material both affect the perception of translucency (Chadwick et al., 2018); *maximum likelihood conjoint measurement* (MLCM) can be applied to pairwise comparison judgements to quantify the

interaction between the multiple physical continua to the perceptual scale (Ho et al., 2008; Knoblauch & Maloney, 2012, Chapter 8).

In both threshold and suprathreshold experiments, reaction time measurements can also be fed into the analysis of the psychophysical data. Reaction times contain information about how easy or difficult the observer found a particular detection or comparison, with reaction times being quicker when a stimulus is much above threshold, and, thus, the detection or decision is easier to make and slower when closer to threshold. One of the advantages of using reaction times is that they can be applied to almost any stimulus (including tasks probing other neural mechanisms such as working memory or attention). Reaction time analysis depends on the specific experimental design and is not further detailed here.

<h1>**Frontiers of Psychophysics**

<bt>Psychophysics has a long past, and it is here to stay—not least as it is one of the only ways to learn about how we (and other organisms) sense, perceive, and make sense of the world around us. Importantly, the way we do psychophysics is, in many cases strongly constrained by the technology that is available to display stimuli, structure the task and procedure, and analyze data. As a consequence, the use of psychophysical methods is far from stagnant and indeed a field with a high drive for innovation.

In the following sections, we highlight some developments at the frontiers of psychophysics. Our examples are from the fields of vision. We identify a series of “sites of innovation” that are not specific to the visual modality, namely: (a) more precise stimulus control, (b) more realistic, naturalistic, and immersive stimuli, (c) development of input-computable models, and (d) developments towards more open psychophysics (in the “open scholarship” sense).

<h2>***More Precise Stimulus Control***

<bt>Historically, the introduction of computer equipment allowing for the presentation of near-arbitrary stimuli revolutionised the ability to run psychophysical experiments. In the visual domain, the computer display has become a key piece of psychophysical apparatus, allowing for the display of arbitrary stimulus patterns and subsequent collection of responses in an integrated fashion.

We noted earlier that we could use RGB monitors to display a wide range of colors visible to humans because human vision is trichromatic. While conventional three-primary-color monitors only allow for the precise production of the signals of three photoreceptor classes, matched to the number of cone classes in the retina, we now know that in addition to the low-light sensitive rods, there is another class of photoreceptors in the human retina. The intrinsically photosensitive retinal ganglion cells (ipRGCs), which are known to be involved in driving nonvisual responses to light such as circadian rhythms, sleep-wake control, mood, pupil size, and alertness, express the short-wavelength-sensitive photopigment melanopsin and encode environmental light intensity independent of the cones and rods (but also receive synaptic input from cones and rods) (Do, 2019; Lucas et al., 2014; Provencio et al., 2000; Spitschan, 2019). Conventional three-primary color displays do not allow for the reproduction and control of rod and ipRGC signals, requiring at least five photoreceptor classes (Hexley, et al., 2021). Generating stimuli that stimulate only one photoreceptor class (while producing no signal in another set of classes) is done using the method of silent substitution (Estévez & Spekreijse, 1982; Spitschan & Woelders, 2018). In practice, this often requires the development of custom stimulation systems or displays (Bayer et al., 2015; Cao et al., 2015; Hexley et al., 2020; Pokorny et al. 2004).

Over the past 30 years, technology has been developed and become available to image the retina at high resolution, resolving single photoreceptors at the cellular level (Roorda & Williams, 1999). Referred to as “adaptive optics,” these techniques allow us to compensate for the optical imperfections of the eye that would otherwise render the imaging of the cones and rods at single-cell resolution impossible. More recently, adaptive optics retinal imaging has been combined with psychophysical techniques to deliver stimuli to single photoreceptors (Harmening & Sincich, 2019). This allows for examination of the sensitivity of single cones (Bruce et al., 2015), spatial integration (Schmidt et al., 2018), and their contributions to the perception of colur (Harmening et al., 2014; Tuten et al., 2017).

<h2>***More Realistic, Naturalistic and Immersive Stimuli***

<bt>The increase in computing power with time has made realistic rendering of visual stimuli, along with the ability to parametrize them, widely accessible. Photorealistic rendering engines allow for the simulation of optical interactions between and within objects under complex illumination, producing compelling images that can be used in psychophysical experiments. In the visual domain, the use of photorealistic stimuli that can be manipulated and parametrized is key to understanding aspects of high-level perception. There are many examples where computer graphics have been used to provide insight into our perception of materials (e.g., translucency, Fleming & Bülthoff, 2005). While the use of naturalistic stimuli allows for generating data that are more ecologically valid, their use may limit the ability to build models “from parts” that can be synthetically modulated (Rust & Movshon, 2005). Importantly, naturalistic stimuli require means to capture naturally occurring scenes in a calibrated fashion (see Morimoto et al., 2019, for an approach for capturing hyperspectral light fields).

Parallel and in lockstep with the development of more advanced computer graphics

techniques, virtual reality (VR) and augmented reality (AR) techniques now enable the presentation of stimuli in a more realistic fashion. Head-mounted VR displays have been on the consumer market for a while and have proven to be an invaluable tool for research. Indeed, VR is one area in which psychophysical methods indeed help with the development of novel displays (Scarfe & Glennerster, 2019). Similarly, AR allows for the mixing of on-the-fly parametric stimulus control with concurrent viewing of the natural world (Hassani & Murdoch, 2019; Murdoch, 2020).

<h2>***Towards Input-Computable Models***

<bt>A core component of psychophysics is modelling human data using various classes of psychophysical models of varying structure and nature. These models go well beyond simply analyzing the data but seek to provide evidence for the computations that the brain performs in solving a specific psychophysical task to gain mechanistic understanding (Pugh & Andersen, 2008). In vision, modelling psychophysical data has a long history. A contemporary approach to modelling of psychophysical data is the development of image-computable models. Image-computable models require an input image—such as a stimulus seen by the visual system—and produce an output. This is encapsulated in the phrase “pixels in, estimates out” (Burge, 2020). For low-level vision, the ISETBio model framework (Cottaris et al., 2019) enables the modelling of retinal signals based on spectral radiance images. Finally, artificial neural network models can be used to constrain psychophysical models (Saxe et al., 2021).

*<h2>**More Open Psychophysics***

<bt>The use of open-source software in psychophysics has a long history, with some core software to run psychophysical experiments, such as the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) and PsychoPy (J. Peirce et al., 2019; J. W. Peirce, 2007, 2008), being

available as an open-source tool from the start. More recently, principles of open scholarship have been adopted in psychology, a field which has been said to be in a “reproducibility crisis.” As many journals and funders now require data to be open and available, the sharing of psychophysical data will also become more common, thereby facilitating meta-analytic approaches to synthesize evidence (see Baker et al., 2018 for an example).

<h1>**Conclusion**

<bt>Psychophysics has a large history, and psychophysical measurements continue to be an integral part of sensory and perception science. Recent developments in stimulus control and how experiments are run have provided innovative impulses in the study of how we see and perceive the world around us.

<refhhbk>**References**

<ref>Baker, D. H., Lygo, F. A., Meese, T. S., & Georgeson, M. A. (2018). Binocular summation revisited: Beyond √2. *Psychological Bulletin*, *144*(11), 1186–1199. https://doi.org/10.1037/bul0000163

Bayer, F. S., Paulun, V. C., Weiss, D., & Gegenfurtner, K. R. (2015). A tetrachromatic display for the spatiotemporal control of rod and cone stimulation. *Journal of Vision*, *15*(11), 15. https://doi.org/10.1167/15.11.15

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. https://www.ncbi.nlm.nih.gov/pubmed/9176952. https://doi.org/10.1163/156856897X00357

Brainard, D. H., Pelli, D., & Robson, T. (2002). Display characterization. In J. Hornak (Ed.), *Encyclopedia of imaging science and technology* (pp. 172–188). Wiley. https://doi.org/10.1002/0471443395.img011

Bruce, K. S., Harmening, W. M., Langston, B. R., Tuten, W. S., Roorda, A., & Sincich, L. C. (2015). Normal perceptual sensitivity arising from weakly reflective cone photoreceptors. *Investigative Ophthalmology & Visual Science*, *56*(8), 4431–4438. https://doi.org/10.1167/iovs.15-16547

Burge, J. (2020). Image-computable ideal observers for tasks with natural stimuli. *Annual Review of Vision Science*, *6*(1), 491–517. https://doi.org/10.1146/annurev-vision-030320-041134

Cao, D., Nicandro, N., & Barrionuevo, P. A. (2015). A five-primary photostimulator suitable for studying intrinsically photosensitive retinal ganglion cell functions in humans. *Journal of Vision*, *15*(1), 15 11 27. doi:10.1167/15.1.27

Chadwick, A. C., Cox, G., Smithson, H. E., & Kentridge, R. W. (2018). Beyond scattering and

absorption: Perceptual unmixing of translucent liquids. *Journal of Vision*, *18*(11), 18. https://doi.org/10.1167/18.11.18

Cottaris, N. P., Jiang, H., Ding, X., Wandell, B. A., & Brainard, D. H. (2019). A computational-observer model of spatial contrast sensitivity: Effects of wave-front-based optics, cone-mosaic structure, and inference engine. *Journal of Vision*, *19*(4), 8. https://doi.org/10.1167/19.4.8

Do, M. T. H. (2019). Melanopsin and the Intrinsically Photosensitive Retinal Ganglion Cells: Biophysics to Behavior. *Neuron*, *104*(2), 205–226. https://doi.org/10.1016/j.neuron.2019.07.016

Estévez, O., & Spekreijse, H. (1982). The “silent substitution” method in visual research. *Vision Research*, *22*(6), 681–691. https://doi.org/10.1016/0042-6989(82)90104-3

Fleming, R. W., & Bülthoff, H. H. (2005). Low-Level Image Cues in the Perception of Translucent Materials. *ACM Transactions on Applied Perception*, *2*(3), 346–382. https://doi.org/10.1145/1077399.1077409

Fort, K., Adda, G., & Cohen, K. B. (2011). Amazon Mechanical Turk: Gold Mine or Coal Mine? *Computational Linguistics*, *37*(2), 413–420. https://doi.org/10.1162/COLI_a_00057

Frisby, J. P. (1980). *Seeing: Illusion, brain, and mind.* Oxford University Press.

Gescheider, G. A. (1997). *Psychophysics : the fundamentals* (3rd ed.). L. Erlbaum Associates.

Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics*. R. E. Krieger Pub. Co.

Harmening, W. M., & Sincich, L. C. (2019). Adaptive optics for photoreceptor-targeted psychophysics. In J. Bille (Ed.), *High Resolution Imaging in Microscopy and Ophthalmology* (pp. 359–375). Springer. https://doi.org/10.1007/978-3-030-16638-0_17

Harmening, W. M., Tuten, W. S., Roorda, A., & Sincich, L. C. (2014). Mapping the perceptual grain of the human retina. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(16), 5667–5677. https://doi.org/10.1523/JNEUROSCI.5191-13.2014

Hassani, N., & Murdoch, M. J. (2019). Investigating color appearance in optical see-through augmented reality. *Color Research and Application*, *44*(4), 492–507. https://doi.org/10.1002/col.22380

Hexley, A. C., Morimoto, T., Smithson, H. E., & Spitschan, M. (2021). Beyond colour gamuts: Novel metrics for the reproduction of photoreceptor signals. *bioRxiv*. Advance online publication. 10.1101/2021.02.27.433203

Hexley, A. C., Yöntem, A. O., Spitschan, M., Smithson, H. E., & Mantiuk, R. (2020). Demonstrating a multi-primary high dynamic range display system for vision experiments. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *37*(4), A271–A284. https://doi.org/10.1364/JOSAA.384022

Ho, Y. X., Landy, M. S., & Maloney, L. T. (2008). Conjoint measurement of gloss and surface texture. *Psychological Science*, *19*(2), 196–204. https://doi.org/10.1111/j.1467-9280.2008.02067.x

Johnson, C. A. (2013). Psychophysical factors that have been applied to clinical perimetry. *Vision Research*, *90*, 25–31. https://doi.org/10.1016/j.visres.2013.07.005

Kingdom, F. A. A., & Prins, N. (2016). *Psychophysics: a practical introduction* (2nd ed.) Elsevier/Academic Press.

Knoblauch, K., & Maloney, L. T. (2012). *Modeling Psychophysical Data in R*. Springer. https://doi.org/10.1007/978-1-4614-4475-6

Koenderink, J. J. (1999). Virtual psychophysics. *Perception*, *28*(6), 669–674. https://doi.org/10.1068/p2806ed

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, *39*(16), 2729–2737. https://doi.org/10.1016/S0042-6989(98)00285-5

Krauskopf, J., Williams, D. R., & Heeley, D. W. (1982). Cardinal directions of color space. *Vision Research*, *22*(9), 1123–1131. https://doi.org/10.1016/0042-6989(82)90077-3

Li, Q., Joo, S. J., Yeatman, J. D., & Reinecke, K. (2020). Controlling for participants' viewing distance in large-scale, psychophysical online experiments using a virtual chinrest. *Scientific Reports*, *10*(1), 904. https://doi.org/10.1038/s41598-019-57204-1

Lucas, R. J., Peirson, S. N., Berson, D. M., Brown, T. M., Cooper, H. M., Czeisler, C. A., Figueiro, M. G., Gamlin, P. D., Lockley, S. W., O'Hagan, J. B., Price, L. L., Provencio, I., Skene, D. J., & Brainard, G. C. (2014). Measuring and using light in the melanopsin age. *Trends in Neurosciences*, *37*(1), 1–9. https://doi.org/10.1016/j.tins.2013.10.004

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge University Press.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Lawrence Erlbaum Associates.

Maloney, L. T., & Yang, J. N. (2003). Maximum likelihood difference scaling. *Journal of Vision*, *3*(8), 573–585. https://doi.org/10.1167/3.8.5

Martin, C. (2014). Making colour in the mind. *Lancet Neurology*, *13*(10), 975. Advance online publication. https://doi.org/10.1016/S1474-4422(14)70147-4

Mollon, J. D., Bosten, J. M., Peterzell, D. H., & Webster, M. A. (2017). Individual differences in visual science: What can be learned and what is good experimental practice? *Vision Research*, *141*, 4–15. https://doi.org/10.1016/j.visres.2017.11.001

Morgan, M. J., Melmoth, D., & Solomon, J. A. (2013). Linking hypotheses underlying Class A and Class B methods. *Visual Neuroscience*, *30*(5-6), 197–206. https://doi.org/10.1017/S095252381300045X

Morimoto, T., Kishigami, S., Linhares, J. M. M., Nascimento, S. M. C., & Smithson, H. E. (2019). Hyperspectral environmental illumination maps: Characterizing directional spectral variation in natural environments. *Optics Express*, *27*(22), 32277–32293. https://doi.org/10.1364/OE.27.032277

Moss, A. J., Rosenzweig, C., Robinson, J., & Litman, L. (2020). Is it ethical to use mechanical Turk for behavioral research? Relevant data from a representative survey of MTurk participants and wages. https://doi.org/10.31234/osf.io/jbc9d

Murdoch, M. J. (2020). Brightness matching in optical see-through augmented reality. *Journal of*

the Optical Society of America. A, Optics, Image Science, and Vision, *37*(12), 1927–1936. https://doi.org/10.1364/JOSAA.398931

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1-2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*, https://doi.org/10.3389/neuro.11.010.2008

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://www.ncbi.nlm.nih.gov/pubmed/9176953. https://doi.org/10.1163/156856897X00366

Pokorny, J., Smithson, H., & Quinlan, J. (2004). Photostimulator allowing independent control of rods and the three cone types. *Visual Neuroscience*, *21*(3), 263–267. https://doi.org/10.1017/S0952523804213207

Prins, N., & Kingdom, F. A. A. (2018). Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes toolbox. *Frontiers in Psychology*, *9*, 1250. https://doi.org/10.3389/fpsyg.2018.01250

Provencio, I., Rodriguez, I. R., Jiang, G., Hayes, W. P., Moreira, E. F., & Rollag, M. D. (2000). A novel human opsin in the inner retina. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *20*(2), 600–605. https://www.ncbi.nlm.nih.gov/pubmed/10632589. https://doi.org/10.1523/JNEUROSCI.20-02-00600.2000

Pugh, E. N., Jr., & Andersen, O. S. (2008). Models and mechanistic insight. *The Journal of General Physiology*, *131*(6), 515–519. https://doi.org/10.1085/jgp.200810041

Read, J. C. (2015). The place of human psychophysics in modern neuroscience. *Neuroscience*, *296*, 116–129. https://doi.org/10.1016/j.neuroscience.2014.05.036

Roorda, A., & Williams, D. R. (1999). The arrangement of the three cone classes in the living human eye. *Nature*, *397*(6719), 520–522. https://doi.org/10.1038/17383

Rust, N. C., & Movshon, J. A. (2005). In praise of artifice. *Nature Neuroscience*, *8*(12), 1647–1650. https://doi.org/10.1038/nn1606

Sauter, M., Draschkow, D., & Mack, W. (2020). Building, hosting and recruiting: A brief introduction to running behavioral experiments online. *Brain Sciences*, *10*(4), 251. https://doi.org/10.3390/brainsci10040251

Saxe, A., Nelli, S., & Summerfield, C. (2021). If deep learning is the answer, what is the question? *Nature Reviews Neuroscience*, *22*(1), 55–67. https://doi.org/10.1038/s41583-020-00395-8

Scarfe, P., & Glennerster, A. (2019). The science behind virtual reality displays. *Annual Review of Vision Science*, *5*(1), 529–547. https://doi.org/10.1146/annurev-vision-091718-014942

Schmidt, B. P., Sabesan, R., Tuten, W. S., Neitz, J., & Roorda, A. (2018). Sensations from a

single M-cone depend on the activity of surrounding S-cones. *Scientific Reports*, *8*(1), 8561. https://doi.org/10.1038/s41598-018-26754-1

Schneider, B. (1980). Individual loudness functions determined from direct comparisons of loudness intervals. *Perception & Psychophysics*, *28*(6), 493–503. https://doi.org/10.3758/BF03198817

Schneider, B., Parker, S., & Stein, D. (1974). The measurement of loudness using direct comparisons of sensory intervals. *Journal of Mathematical Psychology*, *11*(3), 259–273. https://doi.org/10.1016/0022-2496(74)90022-4

Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design. *Psychonomic Bulletin & Review*, *25*(6), 2083–2101. https://doi.org/10.3758/s13423-018-1451-8

Spitschan, M. (2019). Melanopsin contributions to non-visual and visual function. *Current Opinion in Behavioral Sciences*, *30*, 67–72. https://doi.org/10.1016/j.cobeha.2019.06.004

Spitschan, M., & Woelders, T. (2018). The method of silent substitution for examining melanopsin contributions to pupil control. *Frontiers in Neurology*, *9*, 941. https://doi.org/10.3389/fneur.2018.00941

Todor, J. I. (1975). Changes in depth perception during a non-stop walk of 302 1/4 miles. *Perceptual and Motor Skills*, *40*(3), 762. https://doi.org/10.2466/pms.1975.40.3.762

Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research*, *35*(17), 2503–2522. https://doi.org/10.1016/0042-6989(95)00016-X

Treutwein, B., & Strasburger, H. (1999). Fitting the psychometric function. *Perception & Psychophysics*, *61*(1), 87–106. https://doi.org/10.3758/BF03211951

Tuten, W. S., Harmening, W. M., Sabesan, R., Roorda, A., & Sincich, L. C. (2017). Spatiochromatic interactions between individual cone photoreceptors in the human retina. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(39), 9498–9509. https://doi.org/10.1523/JNEUROSCI.0529-17.2017

Wade, N. J. (1998). *A natural history of vision*. MIT Press.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2), 113–120. https://doi.org/10.3758/BF03202828

Webster, M. A. (2015). Visual adaptation. *Annual Review of Vision Science*, *1*(1), 547–567. https://doi.org/10.1146/annurev-vision-082114-035509

Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*(8), 1293–1313. https://doi.org/10.3758/BF03194544

Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, *63*(8), 1314–1329. https://doi.org/10.3758/BF03194545

Willmann, G., Ivanov, I. V., Fischer, M. D., Lahiri, S., Pokharel, R. K., Werner, A., & Khurana, T. S. (2010). Effects on colour discrimination during long term exposure to high altitudes on Mt Everest. *The British Journal of Ophthalmology*, *94*(10), 1393–1397. https://doi.org/10.1136/bjo.2009.178491

Xiao, K., Fu, C., Karatzas, D., & Wuerger, S. (2011). Visual gamma correction for LCD displays. *Displays*, *32*(1), 17–23. https://doi.org/10.1016/j.displa.2010.09.003

<bxn>**Box 24.1.** <bxt>**Classical Psychophysical Laws**

<bxtxt>Weber's law expresses the relationship between background intensity (I) and the minimum intensity change that can be detected (the just-noticeable difference [JND]) (ΔI) mathematically as

$$\Delta I / I = k$$

where k is some constant dependent on the specific modality. Fechner's law expands on Weber's law by relating the sensation, φ, to the stimulus magnitude, S, through the following logarithmic relationship:

$$\varphi = k \log (S)$$

where again, the constant k is dependent on the modality. The mathematically savvy reader may notice that Weber's law is simply the derivative of Fechner's law. Thus the two laws are really one and the same and often referred to collectively as the *Weber–Fechner law*. The Weber–Fechner law is a key principle in psychophysics as it allows us to link the physical stimulus magnitude to perceptual sensation.

At suprathreshold levels, Steven's power law is used instead to equate the magnitude of the perceived sensation, φ, to the magnitude of the physical stimuli, S, as such

$$\varphi = kS^a$$

where k and a are both constants dependent on the specific modality.

<bxn>**Box 24.2.** <bxt>**Signal Detection Theory**

<bxtxt>Signal detection theory is the model underlying the perceptual decision-making

process, that is, what drives an observer to report “seeing” a stimulus on a yes–no task. Perceptual mechanisms are inherently noisy, hence, the sigmoid nature of the psychometric function. Let’s unpack that: The sigmoid shape of the psychometric function tells us that there are some stimulus intensities for which the observer detects somewhere between 0% and 100% of the stimuli presentations. For example, if an observer is asked to report whether they saw a flash of light at a certain stimulus intensity, they will report seeing the flash on only half of the occasions the flash was presented. This is due to the inherent noise in the visual system. Perhaps the observer was capable of seeing the flash on all of those instances but simply blinked or looked away and missed the flash. Or perhaps the observer was anticipating the flash and so reported seeing it when, in fact, nothing was presented. Signal detection theory provides a framework for dealing with the noisy perceptual signals that drive perceptual decisions and, thus, performance in psychophysical tasks.

We can model perceptual mechanisms as having some internal perceptual response that exists in the absence of any external stimulus of a particular modality; we called this the *internal noise distribution.* We model this noise as a Gaussian probability density function, with a mean, μ, and a standard deviation, σ. We can also model the internal perceptual response that exists in the presence of an external stimulus as a Gaussian probability density function with the same standard deviation as the noise but a higher σ and with a higher mean response μs + n**.** We call this the *signal-plus-noise distribution.*

These two distributions inevitably overlap. We can quantify the separation between them using the *sensitivity index, d′* (pronounced dee-prime), defined as

$$d' = \frac{(\mu_{Signal+Noise} - \mu_{Noise})}{\sigma}$$

d′ is simply a standardized measure of our stimulus intensity: the separation between the

signal-plus-noise and noise reported in terms of the standard deviation of the internal noise distributions. As the stimulus intensity increases, *d'* increases, and correspondingly so too does the observer's likelihood to report seeing a stimulus when it is shown. Thus, the signal detection theory model explains why the psychometric function has a sigmoid shape—it is simply the shape of the Gaussian cumulative density function.

One can see that the overlapping Gaussian model leads to a central issue in signal detection theory: Where does the observer set their criteria for detection, and how can we, the experimenter, tell? The arbitrary placement of criterion by the observer is known as *response bias*. On a yes–no task, one can use the *hit rate* (the proportion of correct detections on a trial where the stimulus is presented) and the *false alarm rate* (the proportion of detections on a trial where the stimulus is absent) to calculate *d'* and response bias, *c*, independently from each other as

$$d' = z(Hit\,rate) - z(False\,alarm\,rate)$$

$$c = \frac{z(Hit\,rate) + z(False\,alarm\,rate)}{2}$$

These equations (for derivations, see Green & Swets, 1974; Macmillan & Creelman, 1991, 2005) only hold if the *equal-variance assumption* holds: that is, if the variance of the signal-plus-noise distribution and noise distribution really can be modelled as the same, as we assumed in our signal detection theory model. To check whether the equal variance assumption holds, one can plot a *receiver operating characteristics* (ROC) *curve*, which plots hit rate against false alarm rate. When the equal variance assumption holds, ROC curves will be symmetrical. An asymmetrical ROC tells us the equal variance assumptions fails, and sensitivity cannot be easily disentangled from response bias in this case. To vary the hit rate and false alarm rate in practice, one can ask the observer to give a confidence rating alongside their yes–no response in

a yes–no task, with each confidence level having its own hit rate and false alarm rate that can be plotted along the ROC curve. Alternatively, one could vary the payoffs for the observer, to encourage more liberal observers to become more conservative and vice versa. It is worth emphasizing that alternative force choice tasks are not susceptible to response bias by design and are, thus, often preferred to yes–no tasks, which require an additional response bias analysis as described above.

<bxn>**Box 24.3. Psychophysics from A–Z**

Note: In the below, specific concepts related to the Signal Detection Theory framework are marked by (SDT) in parentheses.

<bxtxt>**Alternative forced choice** (AFC). An alternative forced choice is a bias-free task where the observers are presented with more than two intervals. One of the intervals includes a stimulus, and the observer must respond in which interval the stimulus was presented. The expression *2AFC*, in which stimulus is presented in one of two intervals, is commonly used.

Absolute threshold. Minimum level of stimulus intensity required for detection, for example, the minimum number of photons required for seeing or the minimum amplitude of an auditory stimulus required for hearing.

Adaptation. A neural process to alter the sensitivity of a sensory system after exposure to a certain physical stimulus.

Contrast. The ratio between lowest and highest stimulus magnitude along a specific physical dimension. For example, lightness contrast is often expressed as a ratio between the darkest and brightest light levels.

Criterion (SDT). An observer’s internal threshold of whether or not to respond “yes” on

a given trial. This is often placed by the observer unconsciously and may change as they first start out on a new task before settling.

***d′* (SDT)**. In SDT the separation between the signal-plus-noise and noise distributions reported in terms of the standard deviation of the internal noise distribution $d' = \mu s + n - \mu n / \sigma$.

Category scaling. A method to rate a suprathreshold stimulus by choosing one of the fixed number of categories on which the stimulus falls.

Difference scaling. A method to rate the perceptual difference of stimulus pairs. In a difference scaling task, participants are presented a pair of stimulus pairs and asked to select which pair has a larger perceptual difference.

Discrimination threshold. Minimum difference in intensity between two stimuli along a given physical dimension required for the observer to be able to identify the two stimuli as different from one another, for example, the minimum number of photons required for one light to be perceived as brighter than a second light.

Equal-variance assumption (SDT). Assumption that noise and noise + signal are distributed as normal distributions with equal variance.

False alarm rate (SDT). On a catch trial when the stimulus is not presented, but the observer responds that they detected the stimulus, the result is a false alarm. The false alarm rate is the proportion of false alarms in the total number of catch trials.

Hit rate (SDT). In a trial when the stimulus is presented and the observer responds correctly, it is referred to as a *hit*. The hit rate is the proportion of hits among the total number of trials in which the stimulus was presented.

Interval judgements. A process to decide the magnitude of difference of a pair of stimuli.

Just noticeable difference (JND). Minimum physical difference of a pair of stimuli that can be perceptually discriminated.

Magnitude estimation. Method to quantify the perceptual strengths of a physical stimulus by directly asking participants to assign numerical estimate to the stimulus.

Maximum likelihood conjoint measurement (MLCM). A suprathreshold probabilistic measurement to estimate perceptual scale that explains how two (or more) dimensions of physical stimuli interact and influence a perceptual judgement. Each trial typically consists of a presentation of a pair of stimulus, and participants are asked to make a comparative judgement (e.g., which is heavier) for all possible stimulus pairs multiple times.

Maximum likelihood different scaling (MLDS). A suprathreshold probabilistic measurement to estimate a perceptual scale along a continuous physical dimension based on perceptual judgements regarding difference.

Measure. Process to quantify the magnitude of a physical or perceptual quantity.

Method. Logistic framework to collect data systematically during an experiment.

Observer. Person who participates in an experiment and provides subjective responses.

Ordinal judgements. A process to decide the order of two stimuli according to some dimension. In psychophysics, this would be perceptual strength.

Ordinal perceptual scale. A scale to describe the ranking of a set of stimulus that vary in physical strengths.

Pairwise comparisons. Method to estimate subjective difference or rank among a set of physical stimuli based on paired comparison. Participants are normally asked to make a comparative judgement for every possible pair of stimuli.

Point of subjective equality (PSE). When discriminating a comparison stimulus from a

standard stimulus, the stimulus magnitude of the comparison stimulus that is judged to be perceptually equal to that of the standard stimulus.

Psychometric function. Describes the relationship between the stimulus strength and the probability of the observer's ability to detect or discriminate the stimulus.

Rank order. An order across more than two stimuli along a given dimension.

Response bias. The variation in criterion settings between observers.

Receiver operating characteristics (ROC) **curve (SDT)**. A plot of hit rate against false alarm rate. An ROC curve can be used to check the equal variance assumption: A symmetrical ROC curve tells us the equal variance assumption holds while an asymmetric ROC curve tells us the equal variance assumption has been violated.

Scaling. A measurement to develop a ruler that describes a mapping between physical stimuli along a continuous physical dimension and associated subjective sensation or perception.

Scaling function. A mathematical function to describe a relationship between physical strengths of a stimulus and corresponding perceptual responses.

Sensitivity. The inverse of the threshold (1/threshold).

Signal detection theory (SDT). A framework for uncovering thresholds in a system with inherent internal noise and dealing with confounds such as response bias.

Stimulus. A physical phenomenon that elicits a neural, sensory, or perceptual response (e.g., sound, light).

Stimulus magnitude. A physical strength of an experimental stimulus. Perceptual output is normally measured over a range of stimulus magnitude.

Stimulus reduction. A process to compress the features of complex stimuli, such as ones observed in the real world, into a set of experimentally tractable parameters.

Suprathreshold. Stimulus that is comfortably above threshold, that is, easily distinguishable.

Task. An action or a judgement that a participant is asked to make in a single experimental trial.

Yes–no task. Task in which the observer responds “yes” or “no” to indicate whether they observed the stimulus or an aspect of it presented in a given trial.

<bxn>**Box 24.4** <bxt>**Further Reading on Psychophysical Theory**

<bxtxt>The primary aim of this chapter is to provide an overview of psychophysics as a tool, and we have, therefore, left out theoretical details. Here, we point the reader to some further reading.

Kingdom and Prins (2016) provide a detailed description of psychophysical methods and theory. The associated Palamedes toolbox (http://www.palamedestoolbox.org/) provides a set of MATLAB functions to implement various useful functions, including fitting psychometric function and adaptive staircase procedures.

Gescheider (1997) wrote a classic text on introducing the fundamentals of classical psychophysical methods.

Green and Swets (1974) provide an authoritative overview of signal detection theory.

Morgan, Melmoth, and Solomon (2013) provide a general discussion regarding class A observation applied to threshold-level stimuli and Class B observation for suprathreshold-level stimuli.

Read (2015) discusses the place of psychophysics in contemporary neuroscience, with a specific focus on binocular stereopsis.

Treutwein (1995) discusses adaptive methods in psychophysics, and **Treutwein and Strasburger (1999)** discuss fitting psychometric functions.

Wichmann and Hill (2001a) and **Wichmann and Hill (2001b)** dive deep into psychometric function fitting and doing inference on them.