



# Applying Psychophysics to Applied Spatial Cognition Research

Julia Frankenstein<sup>1,2(✉)</sup>, Fabian Kessler<sup>1,2</sup>, and Constantin Rothkopf<sup>1,2,3</sup>

<sup>1</sup> Department of Psychology, Technical University Darmstadt, Darmstadt, Hesse, Germany

`julia.frankenstein@cogsci.tu-darmstadt.de`

<sup>2</sup> Centre for Cognitive Science, Technical University Darmstadt, Darmstadt, Hesse, Germany

<sup>3</sup> Frankfurt Institute for Advanced Studies, Goethe University, Frankfurt, Hesse, Germany

**Abstract.** Applied spatial cognition research uses many different methods [1], but studies involving psychophysical methods remain quite rare. Here, we argue for the usefulness of psychophysical methods in spatial cognition and use them to address an applied research question. In the present study, we were interested in how sensitive humans are in detecting mismatches between map perspective and the first-person-view of an environment, and whether they exhibit biases in the perception of street angles. We demonstrate that Psychophysics is well-suited to approach these research questions, and provide the reader with a step-by-step description of how we developed our psychophysical study. In particular, we report which problems have to be solved when combining psychophysical methods with complex stimuli often encountered in spatial cognition, e.g. involving complex VR environments, and present our solutions for experimental design and data collection.

**Keywords:** Psychophysics · Map reading · Navigation · Decision making · Spatial perception · Spatial cognition · Perspective change

## 1 Introduction

To navigate their environment successfully, humans need to perceive, represent, and remember space, monitor their locomotion, and develop wayfinding strategies. Spatial cognition research is concerned with all these aspects and is a very interdisciplinary field with a wide range of methods [1]. However, many studies - especially in applied fields of spatial cognition - typically rely on methods such as map drawing [2], pointing [3] and route descriptions [4], or involve the monitoring of navigation behavior, e.g. during wayfinding in real [5] or virtual environments [6]. While the research methods mentioned are appropriate to assess

---

J. Frankenstein and F. Kessler—Contributed equally to this work.

© Springer Nature Switzerland AG 2020

J. Škilters et al. (Eds.): Spatial Cognition 2020, LNAI 12162, pp. 196–216, 2020.

[https://doi.org/10.1007/978-3-030-57983-8\\_16](https://doi.org/10.1007/978-3-030-57983-8_16)

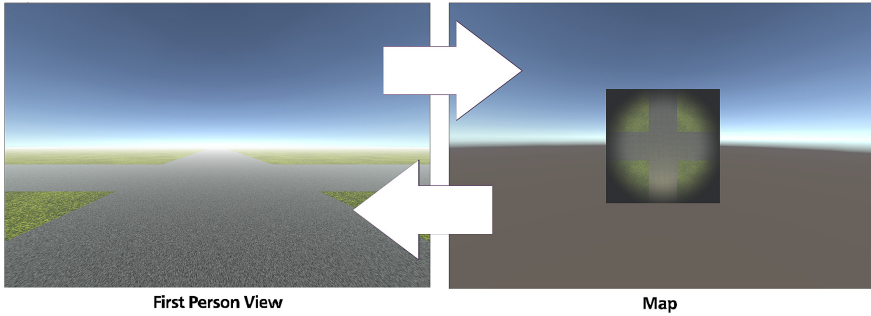
complex human behavior holistically, it often remains difficult to quantitatively describe and explain behavior in terms of the underlying perceptual and cognitive processes. To quantitatively measure cognitive abilities influencing spatial abilities, tests like e.g., the MRT [7] for mental rotation or perspective-taking tests (e.g., [8]) have been developed. These tests often serve as diagnostic procedures, providing a quantitative value describing participants' performance in the respective test dimensions (i.e., a standardized value for participants' ability), and are often used in spatial cognition research to investigate and account for individual differences.

While the development and use of these tests has led to a deeper understanding of the role of individual differences and abilities in spatial behaviour, mostly they are not designed and not appropriate to assess basic perceptual processes or deliver data to develop models of spatial perception and behavior. For that, we think it is useful to apply psychophysical methods. For example, during navigation, we perceive our environment, develop representations of space, and use navigational aids if necessary. The basic perceptual properties of the environment will influence higher cognitive processes, e.g., navigation decisions such as when and where to turn. Therefore, it is necessary to quantify and understand perceptual aspects if we want to develop models of complex spatial behavior. In order to investigate perception of individual factors, psychophysical methods often rely on relatively simple stimuli, compared to the complex VR setups often used in spatial cognition research. Results of psychophysical studies aim to provide insights into the fundamental perceptual processes on which cognitive processes and complex behavior are built on [9].

This paper aims to provide a showcase on how psychophysical methods may be used in and adapted to applied spatial cognition research. To provide an example, we investigated the process of transfer between map perspective and first-person perspective of the environment in an x-crossing scenario (see Fig. 1), i.e., the comparison of a map with the environment, a process necessary e.g., to relocate oneself on a map when being lost [10,11]. To address this question, we used more complex stimuli as compared to classical Psychophysics. We report how psychophysical methods can be adapted to these more complex stimuli, thereby building a bridge between Psychophysics and applied research. Based on preliminary data, we discuss the perspectives and limitations of our methodological approach.

## 1.1 What are Psychophysical Methods?

Psychophysical methods at their core are interested in the quantitative relationship between the physical properties of (environmental) stimuli and their sensory experience [12]. The German physicist G.T. Fechner, one of the pioneering researchers in Psychophysics, during his studies was concerned precisely measuring the differential sensitivity of the different sensory systems in response to varying stimulus intensities [13].



**Fig. 1.** The research question of how sensitive humans are in detecting mismatches between map and environment and whether they exhibit systematic biases in the perception of street angles is an appropriate research question to apply psychophysical methods to. Participants were presented a first-person-view of the environment (left) first, and after that a map view (right), or vice-versa. For details, see Sect. 2.

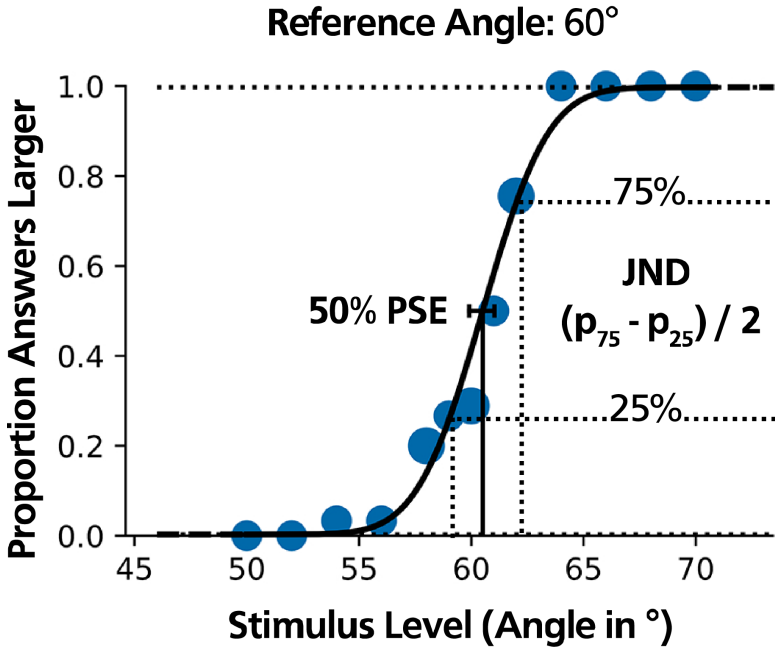
Fechner distinguished two types of sensory thresholds: The absolute threshold, i.e. the threshold when a stimulus can be perceived at all, and the relative threshold, i.e. the minimum difference between two stimuli necessary to discriminate them. The latter is usually referred to as Just-Noticeable-Difference (JND). The point or range, at which two stimuli appear to be equal to the individual participant and could not be discriminated, i.e., participants' performance is not different from chance performance, is called the Point of Subjective Equality (PSE). Both parameters can be derived from the psychometric function, which describes participants' response probabilities at varying stimulus intensities, see Fig. 2. Perfect performance, i.e. no uncertainty about the stimulus, would correspond to a step function, where participants would only resort to guessing when the comparison stimulus and the reference stimulus are the same, but reliably detect differences otherwise.

The precise experimental control and measurement afforded by Psychophysics has a long history of contributing both to theoretical development and understanding of Neurophysiology [14], and serve as a basis for computational models of human perception and information processing [9]. Consequently, the initial ideas by Fechner [13] have been further developed, and adapted to broader questions in many research fields, resulting in numerous psychophysical methods (for an overview, see e.g., [12]). In this paper, we focus on one method to determine differential sensitivity and do not assess e.g., judgements about absolute stimulus magnitude [15].

## 1.2 Psychophysics in Other Disciplines

Psychophysical methods have been introduced successfully to various research fields, leading to quantitative measurements of perceptual and cognitive phenomena, which were only phenomenologically or theoretically described before [14].

Kopie.pdf



**Fig. 2.** The exemplary psychometric function for our research question: We were interested in the Point of Subjective Equality (i.e., angles on map and in perspective view are perceived as equal) and the Just-Noticeable-Difference (i.e., angles on map and in view are reliably perceived as different). Participants had to decide whether the angle of the comparison stimulus was smaller or larger than that of the reference stimulus. For a given reference stimulus (here a 60° angle), the response pattern depending on varying stimulus levels (here comparison angles) are shown (blue dots; diameter proportional to number of trials). Under a set of mathematical assumptions, one can fit the so-called psychometric curve relating the physical stimulus dimension and participants' answers in the decision task. Based on the psychometric curve, one can then extract the PSE, a measure of bias (i.e., in our case the shift of the PSE away from 60°), and the JND, a measure of uncertainty. (Color figure online)

These methods have also aided in quantifying and modeling perceptual biases in orientation discrimination of simple grating stimuli. Humans exhibit better discrimination ability for cardinal directions, commonly referred to as oblique effect [16]. These biases, which were determined in a psychophysical orientation discrimination task, can be explained in terms of the environmental statistics of local orientations found in natural images [17]. Psychophysics has also been applied towards more complex stimuli, such as human faces, providing insight into which features of the face stimulus were most relevant for human face classification [18]. Further, these methods are also applied outside of research in clinical and diagnostic settings, e.g., during standardized hearing tests to detect

the threshold for absolute hearing [19], or to measure visual acuity in humans via the contrast sensitivity function [20].

## 2 Development of Our Method: Bringing Psychophysics into Applied Spatial Cognition Research

Wichmann and Jäkel [21] formulate ‘Best practices’ for designing psychophysical experiments with an emphasis on visual Psychophysics. We used these as a set of guidelines for designing our stimulus and experimental procedure. In the following, we describe some of the necessary steps for developing an experiment using psychophysical methods, namely:

1. Decide for an appropriate research question or adapt the research question in a way that it is possible to use psychophysical methods.
2. Decide for appropriate stimuli (stimulus design including hardware setup and calibration).
3. Decide for an appropriate task and research design (including experimental procedure).
4. Data analysis (fitting psychometric curves, interpreting parameters).

### 2.1 Appropriate Research Question

Psychophysical methods are tools for detecting thresholds (e.g., classification, sensitivity, and perception). The research question should define exactly what should be measured, e.g., whether to assess the absolute or relative threshold, as this determines test design and task [21]. Furthermore, the researcher should decide for the best psychophysical method based on their research question. Here, we show one example of how psychophysical methods may be adapted and used for spatial cognition research, but could not cover the full range of different methods and variations. While we focus on a basic, classic psychophysics method, for research questions like e.g., stimulus detection (e.g., ‘was there a landmark or not?’), signal-detection-approaches may be more appropriate.

We strongly suggest to test only one factor based on psychophysical methods, as usually many trials are necessary for testing and psychophysical experiments are often very exhausting for participants. However, complex or naturalistic stimuli often vary in more than one dimension, this will influence data quality. Quality of the data further depends on participant’s concentration and motivation, too many variations tested, any additional load by changing tasks or complicated answer formats are very likely to decrease data quality. Therefore, these experiments should be kept very structured, and a clear research question helps to achieve this. Especially in the field of applied research, additional data such as e.g., behavioral measurements, map drawing or questionnaires may be of interest. Those can complement psychophysical methods, and sometimes explain high variance in psychophysical data obtained with stimuli varying in more than one dimension, e.g., by revealing individual strategies or metacognition. Due to

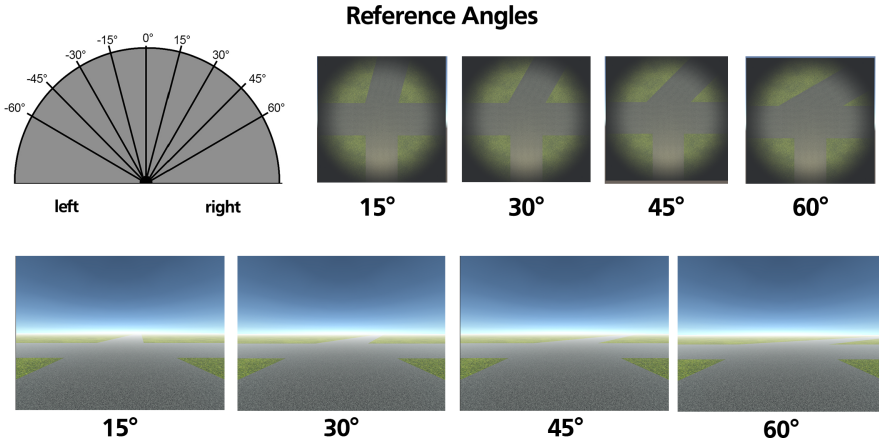
motivational aspects and fatigue, we suggest to keep the use of those additional tasks to a minimum, and schedule them at the end or in an additional session of the experiment.

For our psychophysical study, we were interested in map reading processes. While this is quite a complex behaviour, a basic question is how sensitive humans are in detecting mismatches in street angles between map and environment, and whether there are perceptual biases during perspective change (i.e., during comparison of the map and environment). In psychophysical terms, assessing the street angle sensitivity can be phrased as finding the JND, i.e., the angular difference between the street angles in map and in the environment that can be reliably discriminated by the participants. We also wanted to know whether individuals exhibit systematic biases and differences in sensitivity depending on whether the street deviated to the left or to the right side, and whether the stimulus angle was perceived as larger or smaller than the reference angle. These perceptual biases are quantified by the Point of Subjective Equality (PSE) of the same psychometric curve.

## 2.2 Appropriate Stimuli and Apparatus

To fit an appropriate psychometric curve, many variations of precisely designed and controlled stimuli (e.g., angle sizes, varying colors or sounds) need to be tested. Stimuli should vary in one factor, the stimulus level or intensity, ideally along a continuous ratio-scale (e.g., in our case the street angles). The steps of variation depend on the research question, and could be defined after piloting the material. Computerized methods allow for the precise and objective manipulation of the physical features of the stimuli using a set of parameters [9, 21]. Behavioral responses can be easily captured alongside other quantitative data such as reaction-time and eye-tracking. For spatial cognition research, VR-based technologies are of particular interest. They allow for precise control of stimuli within naturalistic and more ecologically representative settings [22]. Further, they extend stimuli to the spatial domain, and are thus already standard practice in spatial cognition research [6, 23]. However, stimuli for applied research are very likely to vary in more than one dimension, even if generated carefully and with computerized methods. For our study, we designed a simple crossing with one street varying in angle.

However, changing the angle of the street changes not only the angle itself, but in addition the width of the street and the geometry of the center of the crossing (i.e., the space covered by the streets), see Fig. 3. Multi-dimensional variations are a frequent consequence of using naturalistic stimuli. In our case, we still think that our data can be interpreted, as these changes are strongly related to our variation in a geometric function, and are frequently occurring in natural street scenes. By that, these changes are an inherent property of all stimuli we used, and are present (following similar geometric rules) both in maps and in the street scenes. In formulating the research question and interpretation of the data, we strongly advice researchers to be aware of these challenges provided by the stimuli, and to be careful with conclusions focusing on one factor only.



**Fig. 3.** We tested four different reference angles ( $15^\circ$ ,  $30^\circ$ ,  $45^\circ$ ,  $60^\circ$ ) for both the left and right side

**Stimuli.** The virtual crossing environment was developed in Unity (Version 2017.4.2f2), and was comprised of an x-crossing of two roads of equal width (5 virtual meters), surrounded by a wide-open  $500 \times 500$  m grass landscape (see Fig. 3 bottom). For the first-person-view, participants wearing a head-mounted display (HMD) faced the x-crossing from a distance of 9 virtual meters at an eye-height of 1.7 m. The camera was set to a  $60^\circ$  FOV to capture the entire crossing, with a near and far clipping of 0.3 and 1000, respectively. The view in the VR scenery corresponded to participants head movements. The map-view was also presented in the HMD, but the view on the map did not change with participants' head movements. It depicted a top-down rendering of a similar crossing scenario. However, the first-person-view crossing and the map crossing (slightly) differed in one street angle.

The map view was quadratic, but a vignetting was added around the boundaries (see Fig. 3 top). This eliminated the possible use of geometrical cues from the street, coinciding with the sides or corners of the map screen. The map screen in the virtual reality was  $3.5 \text{ m} \times 3.5 \text{ m}$  in size, and positioned 5 m away from the subject. Both inter-trial and inter-stimulus intervals consisted of a grey screen with a luminance matching the mean luminance of the stimulus. By that, the light adaptation process (and fatigue) of participants' eyes was kept to a minimum, and afterimages were prevented. For the same reasons, the inherent contrast of the stimuli was piloted and adapted to be low to prevent fatigue, but still high enough to provide enough information to perform the task reliably. In order to prevent the use of textural cues, we selected textures with low visual inherent orientations and which did not possess large patches. Furthermore, the application of the texture to both the roads and the grass were varied randomly between trials, but color as well as contrasts were kept constant.

**Apparatus.** The crossing and map stimuli (see Fig. 3) were presented in a head-mounted display (Oculus Rift DK II), providing stereoscopic viewing with an  $80^\circ$  (vertical)  $\times$   $80^\circ$  (horizontal) field of view. Further, the HMD had an integrated SMI Eye-tracker (SMI, Teltow, Germany) to track participants' gaze in 3D world-space coordinates throughout the experiment. Using the SMI Unity Plugin, the eye-tracker was calibrated following a 3-point calibration procedure at the beginning of each set of consecutive trials. During the experiment, participants sat on a chair and were able to view the scene freely (i.e. turn their heads). They indicated responses using a wireless Microsoft XBOX 360 controller (Microsoft, Redmond Washington, USA).

### 2.3 Stimulus Presentation, Task, Procedure and Participants

Based on the research question, different psychophysical approaches are appropriate (for an overview, see e.g., [12]). Psychophysical methods may differ in how stimuli are presented, which experimental tasks are performed, and the precise answer format. Due to the nature of stimuli, already established standard experimental tasks, procedures and research designs may not fit to applied research questions without adaptation.

**Stimulus Presentation.** To enable participants to discriminate the comparison from the reference stimulus, these may be presented separated spatially (i.e., both at the same time on one screen, e.g., Two-Alternative-Forced-Choice (2AFC)) or temporally (i.e., one after another, e.g., Two-Interval-Forced-Choice (2IFC)). The minimum presentation time of the stimulus depends not only on the separation format, but is influenced by stimulus complexity. The time must be sufficient for the participant to perceive the stimulus, however, a presentation for too long may not only lead to extended testing time, but may increase involvement of unwanted cognitive strategies.

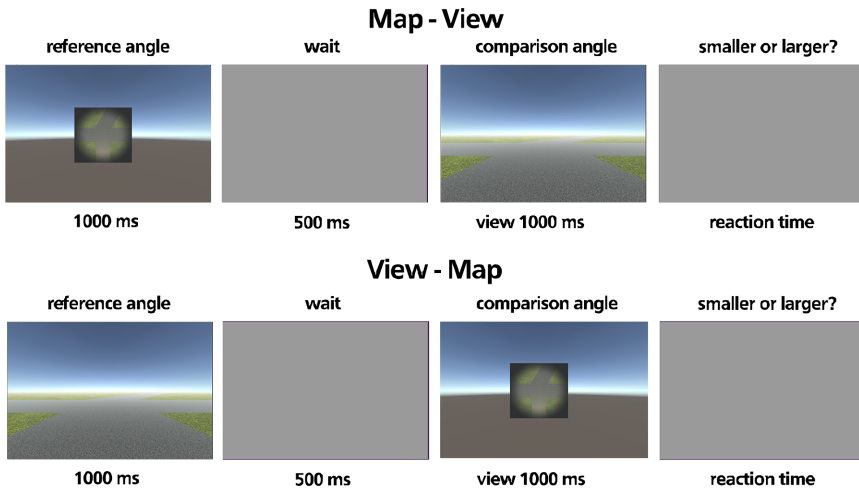
In our Experiment, we adjusted stimulus presentation after comprehensive piloting, and specifically asked our pilot participants for their strategies and their general experience with the Experiment (e.g., afterimages, fatigue). To ensure that presentation time is sufficient, we analyzed pilot participants' performance and fitted psychometrical curves. These served further as basis to decide the number of trials sufficient to fit a full psychometric function [21]. On the basis of piloting, we decided for a presentation time of 1000 ms (see Fig. 4), i.e., participants were not able to perform more than three saccades.

**Task.** Tasks designed must not only match the research question and method, but have to be suitable to be performed by participants over hundreds of trials. Participants may be asked either to (1) identify whether there was a stimulus present during a particular trial (effectively discriminating from zero), (2) whether the presented stimuli were the same or different, (3) which one has the higher value in the dimension of interest (i.e., angle in our study), or (4) which



one is the odd one out among multiple presented stimuli. For an extensive taxonomy of different psychophysical tasks, and for decision criteria of which one is appropriate for certain types of research questions, we refer the reader to [21] and chapter 3 of [24].

For our study, we asked participants to decide whether the depicted street angle in the comparison stimulus presented after the reference stimulus was smaller or larger than the street angle in the reference stimulus. Or equivalently, whether the comparison stimulus was to the left or the right of the reference stimulus. Participants did neither have the opportunity to skip tasks, nor to indicate that both stimuli were equal. To test for the transfer between the two different representations, the reference angle was either shown on the map (Condition: Map - View) or from a first-person view of the crossing (Condition: View - Map).



**Fig. 4.** Psychophysical Task. In every trial, participants wearing a head-mounted display, faced an x-crossing and saw a map of the x-crossing afterwards, or vice versa. The stimulus presented first provided a constant reference angle, and participants had to decide whether the angle of the second stimulus (comparison angle) is smaller or larger.

**General Procedure.** In our study, we used a Two-Interval-Forced-Choice task (2IFC). A single trial consisted of four distinct intervals (three fixed time intervals, 1 variable time interval), two of them containing the reference and comparison stimulus (see Fig. 4). For the first time interval of 1000 ms, participants were presented with a reference angle. In the second interval of 500 ms (inter-stimulus-interval), they viewed at a grey blank screen before facing a second angle, the comparison angle for 1000 ms in the third interval. In the last time interval, participants indicated whether the comparison angle was smaller or larger than the

reference angle, i.e., whether the street in the comparison stimulus was shifted to the left or to the right of the street depicted in the reference stimulus (see Fig. 3). After participants had confirmed their answer, the next trial immediately started. A single trial took approximately 3.25 s to complete. The precise time depended upon subjects reaction time in the fourth interval.

At the beginning of the Experiment, participants received a written instruction, and were allowed to ask questions regarding the procedure to make sure that participants had understood the task. After that, participant completed 10 learning trials (without feedback) to familiarize themselves with the task prior to the first set of measurements. As part of this study we tested four different reference angles ( $15^\circ$ ,  $30^\circ$ ,  $45^\circ$ ,  $60^\circ$ ). Every participant was tested on a single reference angle only (e.g.  $45^\circ$ ), but for both conditions (Map-View and View-Map) and both sides (e.g.,  $15^\circ$  and  $-15^\circ$ ). This yielded 4 psychometric curves for each subject. Trials for one reference angle were presented block-wise, and three blocks were needed to complete measurement for one psychometric curve. Within the three-block measurements, the reference angle was not changed. To account for practise and order effects, we employed a counter-balanced randomized design.

The required number of trials per psychometric curve and the total number of participants depend on the research question, participant's performance, the task and the type of conclusion one wants to draw from the research. For an overview, see [21, 24]. When determining the psychometric function experimentally, the researcher often needs to adapt the range of tested stimuli to the individual participant during the Experiment. By using a suitable method for adaption, adaptive designs make the measurement of the psychometric function more efficient, as more informative stimuli level can be chosen [24].

**Adaptive Testing.** The individual adaptation of stimuli can be achieved by several methods. So-called staircase methods chose appropriate stimuli for trying to converge to participant's PSE based on the responses exhibited on each trial. The different staircase methods differ mostly in how the direction and the appropriate step size in the physical stimulus dimension is chosen [24, 25].

Alternatively, so called 'running fit' methods are based on repeated preliminary fittings of psychometric curves to determine which stimulus intensities to test in subsequent trials. These methods can run on a trial-by-trial basis or between blocks of measurements, but require knowledge about the underlying parametric form of the psychometric function beforehand [24]. In case of the latter, participant's individual stimulus levels can be determined automatically after each trial or manually by the experimenter between sets of measurements.

For our study, we used an experimenter-controlled running-fit method. A full psychometric curve for one angle was measured across three subsequent blocks. After every block, the experimenter fitted a preliminary psychometric curve. By examining a preliminary fit of the psychometric curve and determining those comparison angles which would be most informative about participants' threshold, the stimuli (i.e., the amount and comparison angles) to test in the

next block were specified manually. By that, we ensured that the psychometric curve was fully measured with sufficient data in regions of interest while keeping the number of trials to a minimum. The first and second block consisted of 125 trials each, whereas the third block consisted of 100 trials. Every block lasted for approximately 7–10 min, and participants had a 5 min break between blocks.

Unlike in psychological studies comparing factors, the resulting number of trials and the variation tested are not finally determined beforehand. The number as well as the precise variations (i.e., stimuli) tested may be adapted during data collection to get sufficient data to fit a psychometric curve. During fitting the curve, researchers have to decide when to stop or continue data collection, depending e.g., on the research question, data fit, sample density in the threshold region and the general procedure. We considered data sufficient for fitting the psychometric curve, when (1) the curve exhibited tails for both 0 and 1 in terms of the response probabilities, (2) covered a large range and density of stimuli tested around the threshold and (3) enough data points per stimulus, to have gained sufficient information in the critical region around the threshold (in our case the 0.5 probability, i.e., the PSE). We performed extensive piloting, to identify the approximate number of trials, and tested our blocked design, including the stimuli in the first block, and the effect of subsequent adaption.

In the first block, the angular differences tested were the same for all participants independent of the reference angle, hence there was no participant-specific adaption. The comparison angles in the first block were defined by 5 steps with a step-size of  $2.0^\circ$  in each direction centered around the reference angle. For example, measuring a  $45^\circ$  reference angle comprised the following set of comparison angles: ( $55^\circ, 53^\circ, 51^\circ, 49^\circ, 47^\circ, 45^\circ, 43^\circ, 41^\circ, 39^\circ, 37^\circ, 35^\circ$ ). Within blocks, we presented comparing angles in random order. In psychophysical terms, this is - with respect to the different stimuli shown across trials - known as a non-blocked design [21]. Contrarily, in a blocked design, the same stimulus pair is repeatedly presented for a fixed number of trials.

**Participants.** Classical Psychophysics is known for research designs requiring only a small number of participants but a large number of trials. This rests on the assumption that basic perceptual processes and functions measured are quite similar in the population, therefore, it is sufficient to test a small number of participants only. Ideally, those participants are highly motivated, and reliably attend multiple sessions scheduled on several days, weeks, or months, [21] to perform a vast number of trials. In Psychophysics, it is usually common practice to perform measurements using ‘experienced observers’ [26], unlike in classical psychological Experiments, where Participants normally should be naive about the aim of the Experiment. Providing an intuitive task, clear instructions and practise trials is necessary when testing naive observers.

By testing a low number of participants in many trials, the (statistical) power lies at the individual level. This allows for comparing effects of manipulation within one individual-but less so at the population level-with a high degree of measurement precision. Psychophysical methods were developed to assess

basic functions considered more or less similar in humans. Large quantities of continuous precise measurement data can aid in developing computational and mathematical models to understand the systematic and functional relationships between the physical and psychological variables in those basic functions [27]. For instance, some of the ‘laws’ in psychology were discovered by a single participant and have since then stood the test of time (e.g. Weber-Law or Ebbinghaus’ Forgetting curve).

By contrast, many experiments in both experimental psychology and cognitive science aim at population level parameters and the comparison of different (more or less precisely defined) sub-populations (e.g., young vs. old) or experimental manipulations, often with the goal of performing null-hypothesis significance testing [27].

Applying psychophysical methods to applied spatial cognition research, the number of Participants to be tested will be higher as compared to classical Psychophysics. In many applied scenarios, variance expected may be higher than in testing basic perceptual functions. Furthermore, looking e.g., for individual differences, abilities or learning curves (e.g., cognitive map development), may e.g., require testing of populations or experimental variations.

In our study, data was collected from 16 participants (female = 6) with mean age of 23.88 (SD = 4.06) years. All reported to have either normal or corrected vision, we did not test for visual acuity explicitly. All participants gave their informed consent and received course credit or monetary compensation (8 Euro per hour) for their participation.

## 2.4 Data Analysis

The goal of data analysis is to calculate response probabilities for participants’ behavioral responses at each stimulus intensity obtained during the experiment, fit psychometric curves and extract statistical parameters relevant for answering the research question. As we were interested in participants’ bias and sensitivity when comparing street angles from two different perspectives (i.e., first-person-view and map), we extracted the PSE (reflecting bias) and the JND (reflecting sensitivity) for each angle. We further extracted the reaction times as the length of the ‘smaller or larger?’-decision-interval and eye-tracking data, as they might provide additional insight into the underlying perceptual processes. Because our main focus is on psychophysical methods, we will not describe eye tracking data further in this section.

**Data Preprocessing.** In our study, we collected 24230 trial responses across all participants. Each participant performed a total of approximately 1400 trials distributed across the four conditions (task: map-view/view-map; angle: left/right). Data from the different measurement blocks (for a single angle, direction and task) were combined, after excluding the first ten practice trials. Data were then analyzed for each participant, task, and angle separately, thus yielding four data sets of 350 trials per participant. Responses were coded as either 0

or 1, depending on the street angle of the comparison stimulus was considered to be to the left or to the right of the reference stimulus. We brought all the data in a  $n \times 3$  matrix, where  $n$  refers to the number of different stimulus levels, according to the following format (`stimulus level | nAnswerRight | totalTrials`). This allowed us to calculate response probabilities for each stimulus level (i.e., judgement of comparison angle). For the distribution of responses at each stimulus level, we assumed the beta-binomial mixture model to account for overdispersion in the data, potentially caused by fluctuations in participants performance [28].

**Fitting Psychometric Functions.** Psychometric curves were fitted to participants' responses using the Python version of `psignifit 4` [28], for which also a Matlab version exists. Similar programs are available for R (e.g. `quickpsy` [29]).

The shape of the psychometric function can be chosen according to a Sigmoid family function. All Sigmoid family functions follow a characteristic S-shape, but there exist slight variations (e.g., symmetric/non-symmetric, with or without heavy tails and for stimuli on logarithmic scales), [28]. The choice of which psychometric function to use depends on the stimulus, experimenter's prior intuition based on theory, as well as the scale of the experimental data (e.g., logarithmic vs. non-logarithmic). Additionally, to fit an appropriate psychometric curve, the type of Experiment and/or the tasks (e.g. nAFC, Yes-No-Task or Equal Asymptote) has to be specified.

A total of 4 parameters were required to fit a single curve [21,28]. The threshold  $m$  describes the stimulus intensity at which the psychometric function reaches 0.5 probability level, whereas the width parameter  $w$  describes the scale of the psychometric function, expressed as the difference between stimulus levels between the 0.95 and 0.05 levels. There are two additional parameters  $\lambda$  and  $\gamma$  which describe differences at the asymptote of the Sigmoid (i.e. at 1.0 or 0.0) and correspond to subjects lapse rate (i.e. their probability of giving a false response at extreme stimulus values) and their guess rate (i.e. their probability of randomly guessing). For our study, we chose to model the Sigmoid as a cumulative gaussian, which has the property of being symmetric.

For the parameters of the psychometric function in `psignifit 4` provides Bayesian credible intervals, which allows for assessment different parameters values and comparing them within and across participant. Additionally, researchers can also plot the marginal posterior density for parameter values of the psychometric functions.

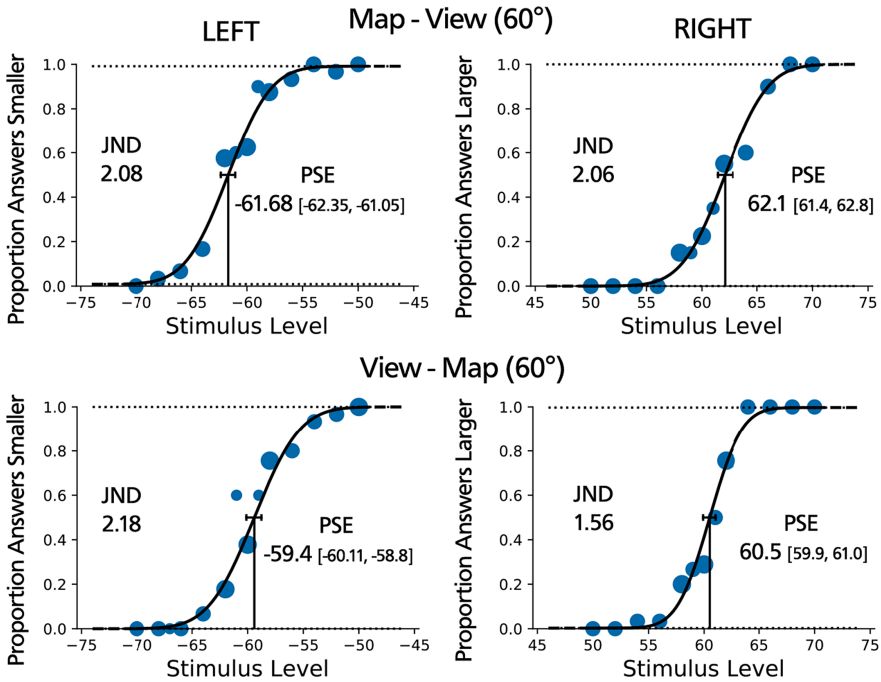
**Extracting Statistical Parameters of Interest.** Once the psychometric curves have been fitted, the researcher can extract statistical parameters, relevant for answering the research question. The fitted psychometric curve allows for obtaining response probabilities along the entire range of the physical stimulus dimension. As stated above and illustrated in Fig. 2, we were interested in the PSE, which in our task can be found at the 50% response probability and coincides with the threshold  $m$  of the psychometric function. We were also interested in the degree of uncertainty, one can either use the width parameter  $w$  of the psychometric function directly, which captures the stimulus range from 0.05 to 0.95 of the response probabilities or one can obtain the stimulus levels at other positions of the psychometric (i.e. 25% and 75%) and calculate the JND as follows:

$$JND = \frac{p_{75} - p_{25}}{2} \quad (1)$$

The fitted psychometric curves, allows for assessment of angular bias and sensitivity for individuals in the four conditions ( $2 \times 2$ : task/direction) under which they performed the task, as well as across individuals on a population level. Our experiments were designed for the former, as we only measured 4 participants for each angle. For illustrative purposes, we provide examples for both types of analyses, to demonstrate how the statistical parameters of interested can be interpreted in light of the research question and what to consider.

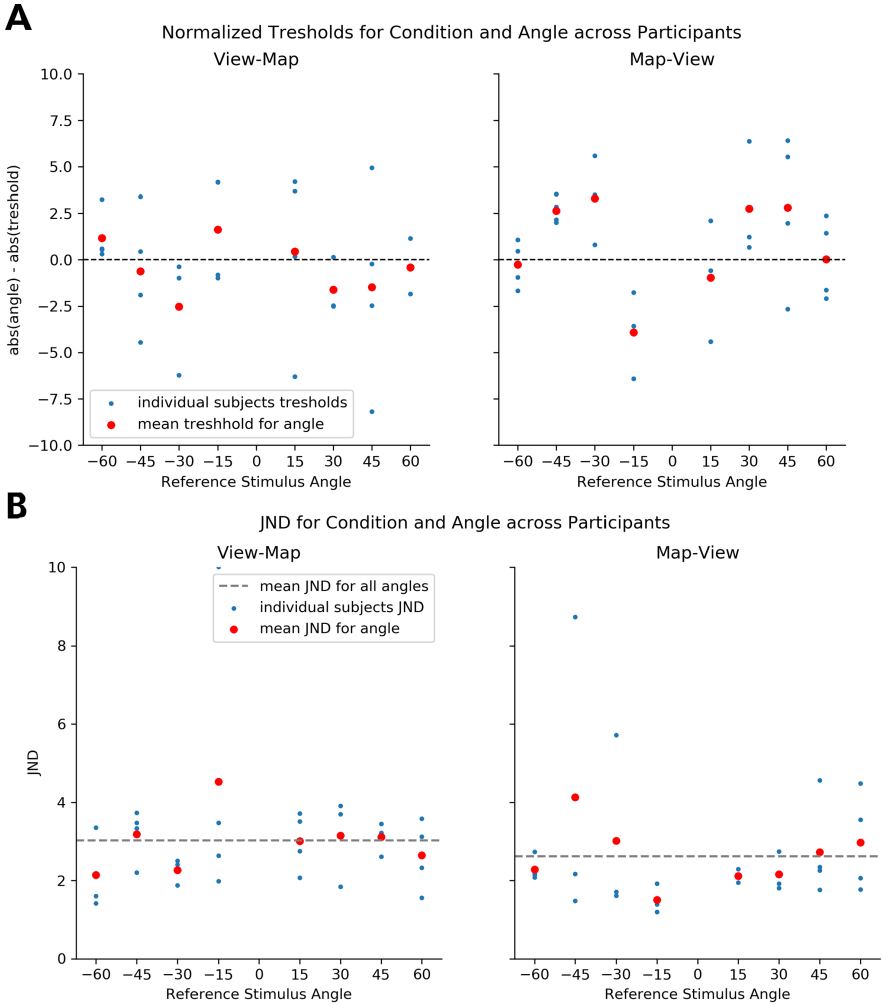
*Analysis for a Single Angle (Individual Participant).* Figure 5 shows an exemplary psychometric curve for a single participant measured at a reference angle of  $60^\circ$  for both procedures (top and bottom), as well as the left and the right side (i.e.  $-60^\circ$  and  $60^\circ$  respectively). Visual inspection of the curves, yields almost unbiased performance in the View-Map condition (PSEs of  $-59.4^\circ$  and  $60.5^\circ$  respectively), there was a slight bias for the Map-View procedure. In the latter, the participant judges the  $60^\circ$  to be larger (average PSEs of  $-61.68^\circ$  and  $62.1^\circ$ ; no overlap of 95% credible intervals with  $60^\circ$ ). Across the two different conditions, there were no systematic differences between the left and the right side for this particular participant, as the 95% credible intervals for the PSE overlap to a substantial degree. Discriminability, which is visually depicted by width of the psychometric curve and quantitatively indicated by the parameter  $w$ , was not significantly different between the two conditions, according to the 95% credible intervals. The JNDs obtained by Eq. 1 indicated that this particular participant could reliably discriminate angles at the 50% level for deviation ranging from  $1.56^\circ$  and  $2.18^\circ$ .

*Analysis Across Angles and Subjects.* To compare the biases of different angles and across all 16 participants, we further normalized our thresholds, effectively centering them around zero (see Fig. 6). Here, we give an overview what types of statements researchers can make from this type of analysis. However, we warn not to jump to conclusions on an insufficient data basis. On the population level, the average normalized PSE varied between the reference angles, i.e., a stimulus deviating from the reference angle between  $-6.4$  and  $6.4^\circ$  (map-view),  $-9.3$ , and



**Fig. 5.** Psychometric curves and extracted parameters for a single participant for a reference angle of 60°. Analysis of one or more participants can be achieved by comparing the 95% Bayesian credibility intervals of the fitted parameters' values. This particular participant exhibits a slight bias for the Map-View condition, judging angles to be larger than 60°, compared to the View-Map condition. There were no systematic differences between the left and right side, within the respective task conditions.

4.9 (view-map) were considered as equal. However, due to a high degree of inter-individual variability in our data and low participant count, those statements about mean thresholds and systematic deviations across the set of all angles might be misleading. Ideally, the design should be adapted towards a tested individual participants across the entire range of angles, to see whether the suggested predictions hold true at the individual level [27]. Visual inspection yields that sensitivity did not vary with the reference angle nor between the left and right sides.



**Fig. 6.** Normalized thresholds and JNDs across subjects, allowing for population based analyses of biases at different angles. Blue dots indicate participants' normalized thresholds/JNDs at a particular angle, red dots indicate mean normalized thresholds/JNDs. We urge experimenters to be careful with respect to interpretation of the result, when only considering a few participants ( $n = 4$  per angle). We plan to test more participants, collecting data of different angles per participant, to perform within-subjects comparisons. Furthermore, we will test more participants in a similar design to facilitate between-subjects comparisons with a larger sample size, and may include additional experimental methods. (A) Participants exhibit biases in both directions (i.e. perceiving some angles as smaller and some angles as larger), their performance is highly heterogeneous. (B) Participants were fairly accurate in correctly discriminating angles. As indicated by the mean JNDs, participants detected mismatches correctly in 50% of the cases, if the stimulus angle deviated from the reference angle at least  $2.4^\circ$  (Map-View) and  $2.7^\circ$  (View-Map). (Color figure online)



### 3 Discussion

In our study, we combined complex stimuli using VR and methods from psychophysics, comparing street angle sensitivity in two different perspectives. This exemplary scenario addresses map reading processes, and perspective changes relevant, e.g., for relocation or navigational decisions during wayfinding when using a map or navigation system. Psychometric curves (see Fig. 2) were fitted to each subject's responses, yielding the PSE (i.e., the bias in comparing angle sizes) and the JND (i.e., the uncertainty in judgment and sensitivity) for each angle. However, due to the low number of participants tested, we consider our results as preliminary. The focus of the paper is to report how to apply psychophysical methods to applied research within the field of spatial cognition. Participants were more sensitive to street angle differences than we expected, based on known difficulties and distortions in developing cognitive maps [30].

Our data show how well participants are able to compare a map view to a first-person-perspective. However, we cannot disambiguate how much of the bias and uncertainty found depend on the change of perspective between the two representations, possibly involving higher cognitive processes, and what can be attributed to the perception of the angles themselves, representing more fundamental perceptual processes. This problem of interpretation may reflect a general problem when using complex setups and stimuli, as e.g., they often vary in more than one factor (i.e., in our case perspective and angle), and increasing stimulus complexity may lead to different individual strategies applied by the participant [9]. E.g., in our study, varying the angle of the street resulted in varying street width at the intersection. To analyze which cue (angle or street width) was used by the individuals, we included eye-tracking. Even though complex stimuli often vary in more than one factor, these influences are present in the physical world as well: Simple, controlled stimuli are rarely present in daily navigation. Nevertheless, we think that psychophysical methods are a valuable tool in spatial cognition research, if - similar to any other research method used - researchers are aware of the limitations in interpreting the data.

Even though further experiments are needed to understand how much of the bias and the uncertainty found are based on perspective change and how much is perceptual, we could detect how sensitive humans are in comparing street angles to a map view and vice versa. These results may influence map design as well as the design of how spatial information is presented in navigation systems. Based on these results, an ideal amount of map schematization could be defined. Too much schematization may be problematic if different shaped crossings look similar. Especially intersections crucial for the navigated route (e.g., crossings requiring a direction change), should be carefully depicted. However, we did not test our participants for their ability in cognitive map development; this will be part of one of our subsequent studies. Furthermore, we aim to include more and more naturalistic stimuli, e.g., intersections of varying geometry (e.g., T-intersections) and increasing environmental complexity (e.g., fog or landmarks up to photorealistic scenes). We want to investigate whether bias or sensitivity changes, depending on the street angle, when other factors, such as the geometry

of the crossing, or visual complexity of the scene, are varied. However, we are aware of the challenges natural stimuli provide. We also plan to test the effect of different eye-heights and participants' relative position in the virtual world on street angle sensitivity.

In the study reported here, we used a classic psychophysical approach to provide a relatively simple example. Alternatively, one could study similar research questions in Spatial Cognition using Signal Detection Theory (SDT), which provides a more comprehensive account for participant's decision making process [31]. SDT takes into account, that participant's responses may not be solely dependent upon the strength of the stimulus (or the magnitude of difference between two stimuli), but that psychological factors may influence the decision as well. E.g., participant's decision criterion can be influenced by experimental design, i.e. the order of trials or the type of instruction. The latter may influence participant's hypotheses about the aim of the experiment or its structure. Participants could also adjust their decision criterion according to their level of motivation, or have hypotheses about the consequences of decisions and errors. These factors, which do not directly depend on the stimulus, are often neglected in classical psychophysical approaches. As psychophysical methods are very diverse, we could not provide a comprehensive guideline within one paper.

We are aware of the fact that psychophysical methods have their limitations. Psychophysical methods are more intricate in assessing higher cognitive processes (e.g., reasoning), individual strategies, spatial knowledge, personal interaction with space, navigation behavior or spatial language. However, we show that in principle psychophysical methods can be combined with relatively complex stimuli to open this method for applied research. However, as psychophysical methods are quite expensive and very demanding for participants, researchers should have a clear research question. Further, they need to decide for an appropriate psychophysical method, design appropriate stimuli and make sure that participants perform the task under the most favorable conditions. This includes providing clear instructions, a sufficient amount of practice trials, the adaptation of stimulus levels towards individual observers to reduce the number of trials, and regular breaks to prevent lapses in attention due to fatigue [21].

Although measurements of psychometric curves are quite expensive, they can provide a valuable tool in spatial cognition research for understanding the influence of different perceptual cues. As perceiving our environment is essential for cognition (i.e., bottom-up processes [32]), investigating and quantifying this basis is in our eyes crucial for a better understanding of higher cognitive processes. This holds true for map reading and perspective change (the example study we depicted in this paper), but could - in principle - be adapted to other basic functions commonly assessed using questionnaires (e.g., mental rotation ability). By that, psychophysical methods may contribute to the development of computerized adaptive test instruments, which quantify participants' ability reliably, providing more precise data than some of the questionnaires in existence. Furthermore, the development of cognitive maps could be better understood, if we knew precisely how and which environmental information is

processed. Psychophysical methods may be useful in quantifying which cues are relevant to understand spatial relations, which cues are selected and integrated, and which information is stored. While reasons for behavior, strategies (e.g., navigation strategies), and memory processes are still hard to assess with these methods alone, a combination of psychophysical methods with other measurements (e.g., spatial tests, map drawing, route descriptions, strategy assessment, behavioural measurements, navigation performance) may contribute to understanding the role of basic functions for these high-level processes. It would be very interesting to what extent psychophysical data predicts performance these other tasks or spatial tests. Using methods from psychophysics within an applied context, investigations under which circumstances (e.g., stress, different tasks, changing stimuli, individual abilities) psychometric curves change may lead to a deeper understanding how basic functions interact with e.g., cognitive functions. These research questions may be addressed by embedding psychometrical measurements in comprehensive experimental designs and use it as a measurement alongside other methods. However, as psychophysical methods require extensive piloting, a lot of research resources and - with more complicated designs and between-subjects-designs - and consequently a lot of reliable participants. Other measurements could directly be combined with psychophysical methods in less complicated designs, e.g., eye tracking, which aids in understanding which spatial information participants look for to solve spatial tasks (especially under time pressure). Based on eye-tracking data, it is possible to investigate when and how regions of visual interest change with varying tasks or stimuli, and how this may influence participants' decisions, captured in the parameters in the psychometric curve. Furthermore, reaction times and confidence ratings could be combined with psychometric measurements of perception, to analyze aspects of the relationship between perception, cognition and meta-cognition.

The focus of our paper was to report how psychophysical methods could be used for applied spatial cognition research. Generally speaking, experiments using these methods need to be based on a profound theoretical motivation, and carefully piloted and designed. We suggested steps to take for developing experiments using classical psychophysical methods and analyzing data in an applied research context.

**Acknowledgements.** We would like to thank Frank Jäkel for his help in designing the experiment, our patient colleagues for their feedback during piloting, and our participants for providing their data.

## References

1. Newcombe, N.: Methods for the study of spatial cognition. *Dev. Spat. Cogn.* **277**, 303–326 (2013)
2. Huynh, N.T., Doherty, S.T.: Digital sketch-map drawing as an instrument to collect data about spatial cognition. *Cartographica* **42**, 285–296 (2007)
3. Frankenstein, J., Mohler, B.J., Bühlhoff, H.H., Meilinger, T.: Is the map in our head oriented north? *Psychol. Sci.* **23**, 120–125 (2012)

4. Hölscher, C., Tenbrink, T., Wiener, J.M.: Would you follow your own route description? Cognitive strategies in urban route planning. *Cognition* **121**, 228–247 (2011)
5. Richardson, A.E., Montello, D.R., Hegarty, M.: Spatial knowledge acquisition from maps and from navigation in real and virtual environments. *Mem. Cogn.* **27**, 741–750 (1999)
6. Diersch, N., Wolbers, T.: The potential of virtual reality for spatial navigation research across the adult lifespan (2019)
7. Vandenberg, S.G., Kuse, A.R.: Mental rotations, a group test of three-dimensional spatial visualization. *Percept. Mot. Skills* **47**, 599–604 (1978)
8. Kozhevnikov, M., Hegarty, M.: A dissociation between object manipulation spatial ability and spatial orientation ability. *Mem. Cogn.* **29**, 745–756 (2001)
9. Waskom, M.L., Okazawa, G., Kiani, R.: Designing and interpreting psychophysical investigations of cognition. *Neuron* **104**, 100–112 (2019)
10. Dai, R., Thomas, A.K., Taylor, H.A.: When to look at maps in navigation: metacognitive control in environment learning. *Cogn. Res.: Princ. Implic.* **3**, 36 (2018)
11. Levine, M., Jankovic, I.N., Palij, M.: Principles of spatial problem solving. *J. Exp. Psychol.: Gen.* **111**(2), 157–175 (1982)
12. Gescheider, G.A.: The classical psychophysical methods. *Psychophys.: Fundam.* (1997)
13. Gibson, J.J., Fechner, G.T.: *Elemente der Psychophysik*, vol. 2. Breitkopf und Härtel, Leipzig (1860)
14. Read, J.C.: The place of human psychophysics in modern neuroscience. *Neuroscience* **296**, 116–129 (2015)
15. Stevens, S.S.: To honor Fechner and repeal his law. *Science* **133**, 80–86 (1961)
16. Appelle, S.: Perception and discrimination as a function of stimulus orientation: the “oblique effect” in man and animals. *Psychol. Bull.* **78**(4), 266 (1972)
17. Girshick, A.R., Landy, M.S., Simoncelli, E.P.: Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat. Neurosci.* **14**, 926–932 (2011)
18. Macke, J.H., Wichmann, F.A.: Estimating predictive stimulus features from psychophysical data: the decision image technique applied to human faces. *J. Vis.* **10**(5), 1–24 (2010)
19. Sanjay, H.S., Bhargavi, S., Madhuri, S.: Auditory psychophysical analysis of healthy individuals based on audiometry and absolute threshold tests. *Int. J. Eng. Technol. (UAE)* **7**, 99–106 (2018)
20. Pelli, D.G., Bex, P.: Measuring contrast sensitivity. *Vis. Res.* **90**, 10–14 (2013)
21. Wichmann, F.A., Jäkel, F.: *Methods in Psychophysics*. Stevens’ Handbook of Experimental Psychology and Cognitive Neuroscience, pp. 1–42. Wiley, Hoboken (2018)
22. Bohil, C.J., Alicea, B., Biocca, F.A.: Virtual reality in neuroscience research and therapy. *Nat. Rev. Neurosci.* **12**, 752–762 (2011)
23. Mallot, H.A., Gillner, S., van Veen, H.A.H.C., Bülthoff, H.H.: Behavioral experiments in spatial cognition using virtual reality. In: Freksa, C., Habel, C., Wender, K.F. (eds.) *Spatial Cognition*. LNCS (LNAI), vol. 1404, pp. 447–467. Springer, Heidelberg (1998). [https://doi.org/10.1007/3-540-69342-4\\_21](https://doi.org/10.1007/3-540-69342-4_21)
24. Kingdom, F.A., Prins, N.: *Psychophysics: A Practical Introduction*, 2nd edn. (2016)
25. Leek, M.R.: Adaptive procedures in psychophysical research. *Percept. Psychophys.* **63**, 1279–1292 (2001)
26. Jäkel, F., Wichmann, F.A.: Spatial four-alternative forced-choice method is the preferred psychophysical method for Naïve observers. *J. Vis.* **6**, 13–13 (2006)

27. Smith, P.L., Little, D.R.: Small is beautiful: in defense of the small-N design. *Psychon. Bull. Rev.* **25**, 2083–2101 (2018)
28. Schütt, H.H., Harmeling, S., Macke, J.H., Wichmann, F.A.: Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vis. Res.* **122**, 105–123 (2016)
29. Linares, D., López-Moliner, J.: quickpsy: an R package to fit psychometric functions for multiple groups. *R J.* **8**, 122–131 (2016)
30. Ishikawa, T., Montello, D.R.: Spatial knowledge acquisition from direct experience in the environment: individual differences in the development of metric knowledge and the integration of separately learned places. *Cogn. Psychol.* **52**, 93–129 (2006)
31. Green, D.M., Swets, J.A.: *Signal Detection Theory and Psychophysics*, vol. 1. Wiley, New York (1966)
32. Pecher, D., Zwaan, R.A.: *Grounding cognition: the role of perception and action in memory, language, and thinking* (2005)