The FechDeck: A Hand Tool for Exploring Psychophysics

JAMES FERWERDA, Rochester Institute of Technology

Learning the methods of psychophysics is an essential part of training for perceptual experimentation, and hands-on experience is vital, but gaining this experience is difficult because good tools for learning are not available. The FechDeck is an ordinary deck of playing cards that has been modified to support learning the methods of psychophysics. Card backs are printed with noise patterns that span a range of densities. Faces are augmented with line segments arranged in "L" patterns. Jokers are printed with ruled faces and with backs that serve as standards. Instructions provided with the FechDeck allow users to perform threshold experiments using Fechner's methods of adjustment, limits, and constant stimuli; scaling experiments using Thurstone's ranking, paired comparison, and successive categories methods; and Stevens's magnitude estimation method. Spreadsheets provided with the deck support easy data entry and meaningful data analysis. An online repository supporting the FechDeck has been established to facilitate dissemination and to encourage open source development of the deck.

CCS Concepts: • **Human-centered computing** → **User models**; **User studies**; **Laboratory experiments**;

Additional Key Words and Phrases: Vision, psychophysics, education, open source

1 INTRODUCTION

Psychophysics is one of the pillars of experimental psychology. Psychophysical methods allow quantitative investigation of the relations between the physical properties of the world and our perceptual experiences. Much of our knowledge about perception and cognition has come from psychophysical experiments, and many important advances in human-centered technologies have been based upon them. However, learning the methods of psychophysics is challenging, because good tools for learning are not available. Examples in books and papers can provide understanding of the concepts but not hands-on experience. Software tools can provide hands-on experience, but systems come and go, and often require students to have significant coding experience to move beyond pre-programmed examples.

This article describes the design and use of the FechDeck, a tool developed to facilitate learning the methods of psychophysics. The FechDeck is an ordinary deck of playing cards that has been modified so that the cards can be used as stimuli in a variety of psychophysical experiments. By manipulating the cards according to instructions

provided with the deck, a user can run a variety of threshold and scaling experiments, and gather data that can be analyzed using spreadsheets, to derive standard psychophysical measures.

In the following sections, we first provide an overview of the field of psychophysics, describing significant developments and their contributors. We then describe the design of the FechDeck and the features that make it useful as a tool for psychophysics education. Next, we give three examples of how the FechDeck can be used to run psychophysical threshold and scaling experiments, and describe the support provided with the deck, which enables analysis of the resulting experimental data. We then briefly describe the online structures we have put in place to make the FechDeck and its resources freely available, and conclude with a discussion of the contributions and limitations of the project, and avenues for future development.

2 BACKGROUND

The history of the field of psychophysics has been well documented elsewhere and will only be briefly reviewed in the following. For detailed accounts, readers should consult Gescheider [1], Lu and Dosher [2], and Kingdom and Prins [3].

The development of modern psychophysical methods is often traced to Ernst Heinrich Weber, a physiologist at the University of Leipzig, Germany, who in the 1830s studied weight perception and found that just-noticeable differences (JNDs) in weight increased in proportion to the magnitude of the weight being evaluated. This finding, which applies in many sensory domains, has been termed *Weber's law* and is expressed as $\Delta I = kI$, where I is the magnitude of a standard stimulus, ΔI is the JND or difference threshold, and k is a fractional constant that varies with sensory domain.

Weber's colleague, Gustav Theodor Fechner, observed the predictions of Weber's law for different stimulus magnitudes and proposed a relation now known as *Fechner's law*, which states that the perceived magnitude of a stimulus can be predicted by the expression $S = klogI$, where $logI$ is the logarithm of the physical stimulus magnitude and k is the fractional constant given by Weber's law.

These two "laws" point to the two fundamental measures provided by psychophysics: thresholds and scales. Thresholds describe the relationship between physical stimulation and sensation at a point in the stimulus domain, and scales predict the relationships between stimulation and sensation over ranges in the stimulus domain.

Fechner went on to develop a variety of experimental methods for measuring sensory thresholds. He termed these the *method of average error*, the *method of just-noticeable differences*, and the *method of right and wrong cases*. These have evolved into the current-day *method of adjustment* (MOA), *method of limits* (MOL), and *method of constant stimuli* (MCS). His publication in 1860 of the book *Elemente der Psychophysik* [4], which describes these methods, is generally regarded as the beginning of the field.

Methods for sensory scaling can also be traced back to Fechner, but his approach was essentially to measure difference thresholds at points along a sensory domain, and then to derive scales using Fechner's law. Scaling methods that do not depend on threshold measurements can be traced to psychologist Louis Leon Thurstone [5], who in the 1920s developed an analytical technique known as the *law of comparative judgment*, which allowed psychophysical scales to be derived from ordinal judgments of paired stimuli. *Paired comparison, ranking, and successive intervals* are the experimental methods most closely associated with Thurstonian scaling. These methods are known as "indirect" or "confusion-based" scaling because they use statistical models to derive interval scales from ordinal judgment tasks, and they require that adjacent stimuli in the stimulus set be close enough that observers sometimes misjudge the stimulus order.

In contrast, the "direct" scaling methods associated with Stevens [6] use simpler statistics and have no constraints on stimulus intervals. The classic direct methods include *fractionation, magnitude production, and magnitude estimation*. In these methods, the observer is either asked to adjust a test stimulus so it stands in some ratio to a standard or is asked to assign a numerical value to a test stimulus that describes its value relative to the value of a given standard. Scale values are found by taking the geometric means of test stimulus values. Across different experimental methods and sensory domains, Stevens found that observers' judgments were well fit by

power functions of the form $S = kI^a$, where I is the physical stimulus magnitude, S is the perceived stimulus magnitude, a is an exponent that depends on the type of stimulus, and k is a constant that relates units in the I and S domains. In many cases, the exponents of Stevens's so-called Power law are less than 1.0, producing compressive functions that yield scales similar to Fechner's law.

One of the constant tensions in psychophysics is the tradeoff between sufficient measurement and observer effort, and many psychophysical experiments are notoriously arduous. With the advent of computer-based testing, a variety of adaptive methods [7] that maximize information gain per experimental trial have been developed. More sophisticated analytical techniques [8–10] have also accompanied these methodological advances, and they allow for more efficient estimation of psychometric functions [11] that describe the relations between stimulus magnitudes and observer responses, and provide the formal link between the physical and psychological realms.

The psychophysical methods described earlier comprise the essential toolkit used by experimental psychologists to understand the capabilities and limitations of sensory systems, and the relations between the physical properties of things in the world and their perceptual appearances. Learning to use these methods is an essential component of training in experimental psychology; however, gaining this experience as a student can be hard, because existing teaching tools, while valuable, also have limitations. Examples in books and chapters, such as those listed previously, can be instructive but do not provide hands-on experience. Software systems [12–15] can provide hands-on experience but are often fragile, and users can face steep learning curves to advance beyond pre-programmed examples (Strasburger [16] provides an excellent guide to these systems).

The FechDeck was imagined as an alternative to existing teaching tools. The motivation was to create a robust, low-cost, widely distributable, enjoyable-to-use, hands-on tool for teaching psychophysics. The following sections describe the design and use of the deck in greater detail.

3 DESIGN

The FechDeck is a standard deck of playing cards, consisting of 54 cards in four suits of 13 cards each and two jokers. In each suit, the card faces run from ace through king and are considered to represent the numbers 1 through 13. The jokers are used as standards in the experiments and represent the numbers 5 and 9. The faces of the cards (Figure 1) are of the conventional "French" style, with the exceptions that L-shaped figures have been added at the borders of numbered cards 2 through 6 and the aces have inset photographs to honor Fechner, Weber, Thurstone, and Stevens. The faces of the jokers show the FechDeck logo and have 10-unit rulers running along each long edge.

The card backs (Figure 2) serve as stimuli in the experiments and are printed with square-pixeled, "salt and pepper" noise patterns with varying percentages of white and black elements. The patterns were generated using the MATLAB image processing toolbox [17]. First, 103×141–pixel binary arrays were created using the "imnoise" function with the "salt" and "pepper" parameter. Then, the arrays were scaled to $824 \times 1{,}128$ pixels using the "imresize" function with "nearest" neighbor interpolation. Finally, the boundary elements in the arrays were removed using the "imcrop" function, and the arrays were quantized and saved as $822 \times 1{,}122$–pixel TIFF images using the "imwrite" function. This process resulted in sharp-edged, bi-level, noise pattern images where each "noise" pixel was represented by an 8×8 array of "image" pixels.

To create the physical deck, the face and back images were then printed on playing cards by a commercial card deck supplier. Each card measures 2.5×3.5 in., so assuming a hand-held viewing distance of 14 inches, each noise image subsumes approximately 10×14 degrees of visual angle, with each noise pixel subsuming approximately 0.1 degrees. Under these conditions, a vertical black/white grating of the same dimensions would have a horizontal spatial frequency of approximately five cycles/degree. The printing and card-coating processes had minimal impact on the clarity of the images, and they appear sharp at normal viewing distances.

Within each suit, the densities of the noise patterns (percentage of black pixels) printed on the card backs increase with card rank. Thus, the aces have the least dense patterns, and the kings have the most dense patterns. The specific densities printed on each card are listed in Table 1. The density ranges for the spades, hearts, and
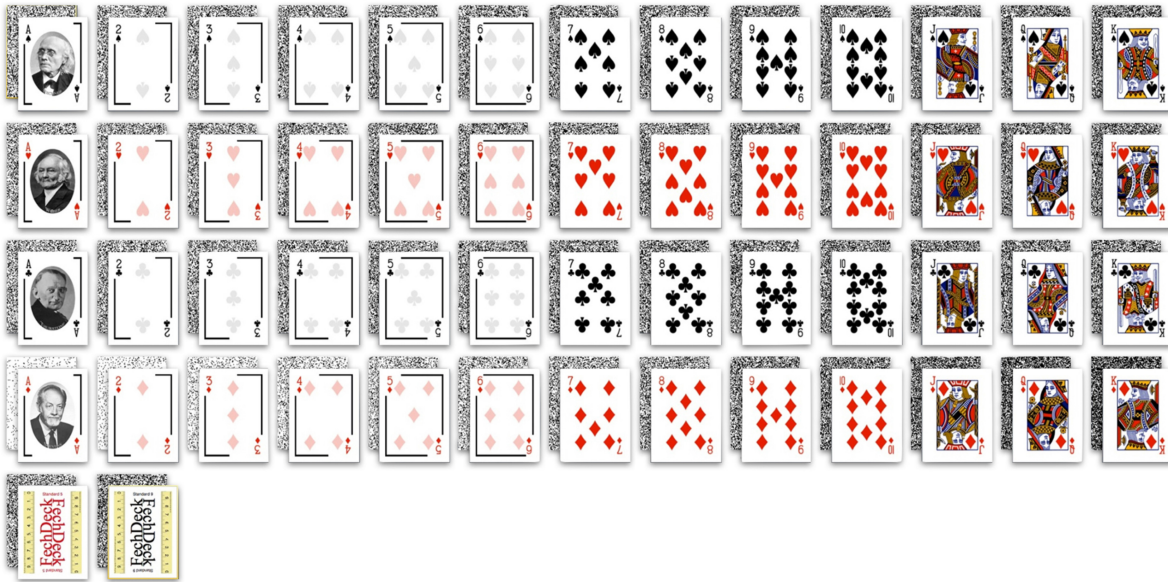
Fig. 1. FechDeck card faces. Faces of the deck are conventional with the exceptions that L-shaped figures have been added to the borders of cards numbered 2 through 6; the aces show photographs that honor the founders of psychophysics, and the jokers show the FechDeck Logo and have rulers running along their long edges.
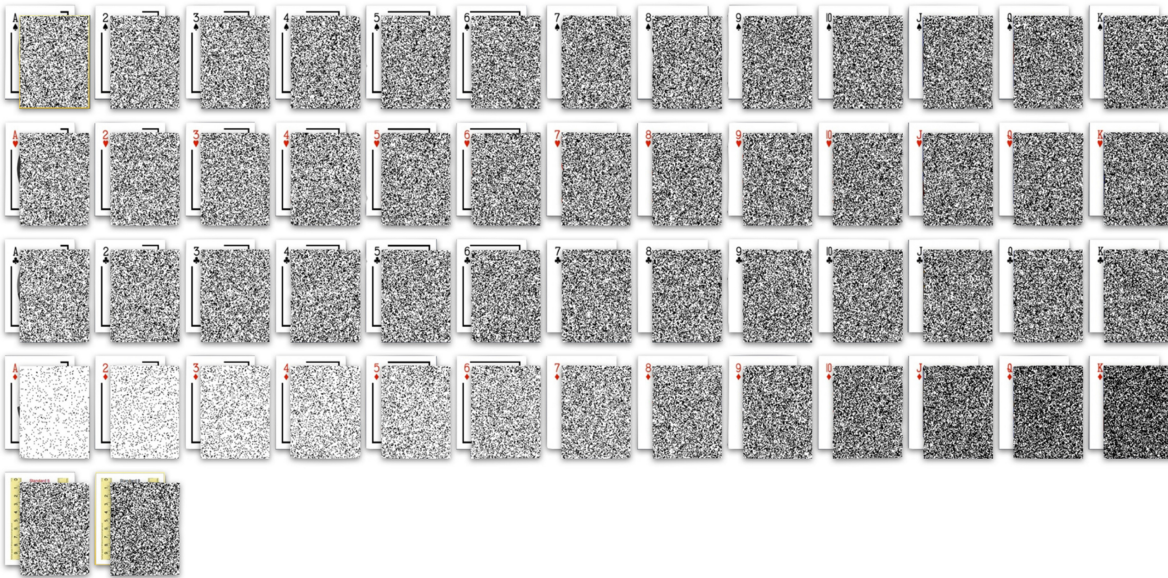
Fig. 2. FechDeck card backs. The card backs are printed with square-pixeled bi-level noise patterns with varying percentages of white and black pixels. These patterns serve as the stimuli for the experiments supported by the FechDeck.

Table 1. Card Noise Pattern Densities

		CARD RANK												
		A	2	3	4	5	6	7	8	9	10	J	Q	K
CARD SUIT	♠	0.3500	0.3625	0.3750	0.3875	0.4000	0.4125	0.4250	0.4375	0.4500	0.4625	0.4750	0.4875	0.5000
	♥	0.3500	0.3625	0.3750	0.3875	0.4000	0.4125	0.4250	0.4375	0.4500	0.4625	0.4750	0.4875	0.5000
	♣	0.3500	0.3625	0.3750	0.3875	0.4000	0.4125	0.4250	0.4375	0.4500	0.4625	0.4750	0.4875	0.5000
	♦	0.0500	0.1000	0.1500	0.2000	0.2500	0.3000	0.3500	0.4000	0.4500	0.5000	0.5500	0.6000	0.6500

Note: Within each suit, the densities of the noise patterns increase with card rank. The spades, hearts, and clubs suits all have the same density range. The diamonds suit has a wider range to support the magnitude estimation method. Although some cards have the same densities, the noise patterns printed on each card are unique.

clubs suits are all 35% to 50%. The density range for the diamonds suit is 5% to 65% to support the wider stimulus range required for magnitude estimation scaling. The "red" joker (numeric 5) has a density of 40%, and the "black" joker (numeric 9) has a density of 45%. Even though the noise patterns densities of some cards are the same, a unique noise pattern was generated for each card.

The ranges and increments of noise densities were selected to provide typical and meaningful results in the supported experimental procedures. The particular levels were initially selected through trial and error by the author, but they have since been confirmed as effective through testing by the author and colleagues with hundreds of observers in both individual and classroom settings. Although the results of individual experiments with particular observers certainly vary, in the aggregate we have found that the results produced by the experiments are typical, meaningful, repeatable, and serve the FechDeck's main purpose of educating about the methods of psychophysics. In this regard, all of the results reported in the following sections are based on real data from actual runs of the FechDeck-supported experiments.

4 USAGE

The FechDeck can be used to run experiments that demonstrate a variety of psychophysical methods. Version 1.0 of the deck includes instructions for running experiments that demonstrate Fechner's three "classical" threshold methods (adjustment, limits, and constant stimuli), as well as both indirect/Thurstonian scaling methods (paired comparison, ranking, successive intervals) and one of Stevens's direct scaling methods (magnitude estimation). The following sections give examples of a selection of these experiments to illustrate how the FechDeck is used in practice.

4.1 Example 1: Method of Constant Stimuli

Of Fechner's three methods for measuring thresholds, the MCS is the most general. On each trial, a test stimulus randomly selected from a discrete stimulus set is compared to a standard, and the observer's task is to indicate if the test has more or less of some attribute than the standard. Over repeated presentations, the frequencies with which each of the test stimuli are judged to be in one category or the other are tabulated for subsequent analysis.

In the first stage of the analysis, the frequency data are converted to probabilities. The relationship between these probabilities and the magnitudes of the test stimuli is known as a psychometric function. In the second stage of the analysis, this empirical psychometric function is modeled with an analytical S-shaped function (typically cumulative normal, logistic, or Weibull (see Wichman and Hill [18] for a tutorial)). Finally, the model is used to estimate threshold measures such as the point of subjective equality (PSE), upper and lower thresholds (UL, LL), and the JND.

The procedure for using the FechDeck to run an experiment with the MCS is illustrated in Figure 3. The instructions are as follows:

1. Place the red joker (standard 5) face down on the table as the "standard."
2. Remove the diamonds suit from the deck. Shuffle the remaining cards to form the "test" pile.
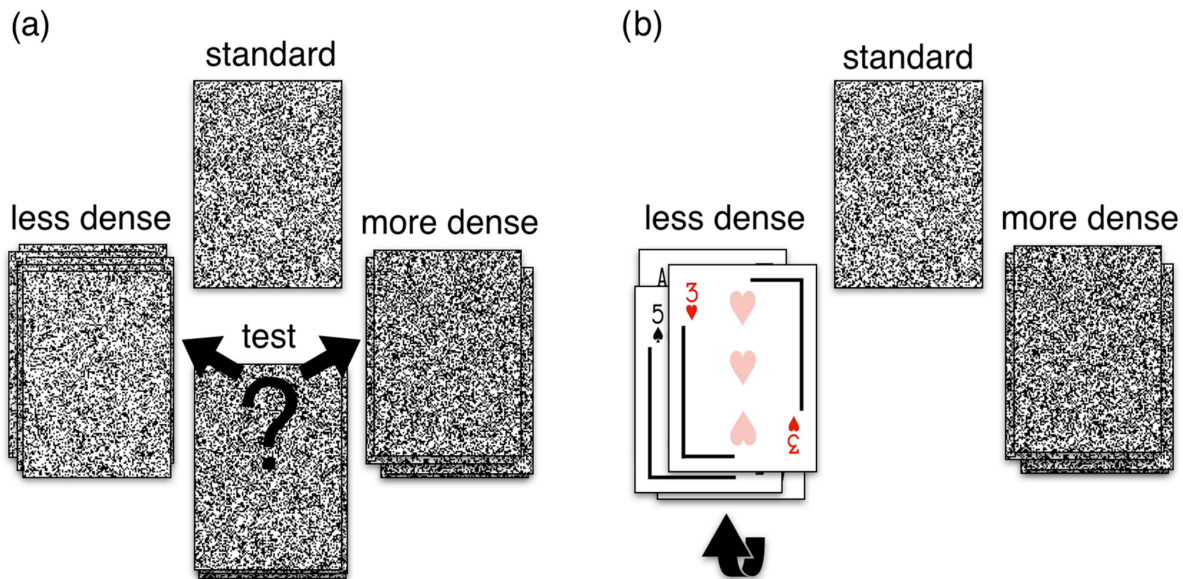
Fig. 3. MCS procedure. (a) Testing. (b) Data gathering.

3. Place the test pile face down below the standard.
4. Starting from the top of the test pile, compare each card to the standard. If the pattern on the card is less dense than the standard, place it on the left, and if it is more dense, place it on the right. Repeat for all the cards in the pile.
5. Take the "less dense" pile, turn it face up, and sort the cards by rank.
6. Enter the number of cards of each rank into the leftmost open "observers/repetitions" column of the "MCS" spreadsheet.
7. Reassemble and shuffle the cards from the less/more dense piles.
8. Observations/repetitions: (Group) Enter each observer's results in separate columns in the spreadsheet; (Individual) Repeat steps 3 through 7 until the desired number of repetitions has been completed.
9. Analyze the data using the spreadsheet.
10. If desired, repeat the procedure using the black joker (standard 9).

An Excel spreadsheet that performs the data analysis for this experiment is included in the FechDeck distribution. The features of this spreadsheet are illustrated in Figure 4. The data show the results from 12 graduate student observers using the FechDeck in a classroom setting.

To perform the analysis, the user first enters the data generated by each repetition of the procedure into the yellow cells in the spreadsheet. The spreadsheet then sums across the rows of these cells to calculate the frequencies with which each test card was judged to be "less dense" than the standard. Subtracting these numbers from the total number of presentations yields the frequencies with which each test card was judged to be "more dense" than the standard. Dividing this number by the total number of presentations converts these numbers to probabilities. The spreadsheet then plots these probabilities versus the card noise densities to visualize the empirical psychometric function (Figure 4(b)).

In the second stage of the analysis, the spreadsheet fits the empirical psychometric function with a cumulative normal model. To derive the parameters of the model, the probabilities are first converted to z-scores. If a cumulative normal model is a good descriptor for the empirical psychometric function, then the z-scores will have a

FechDeck method of constant stimuli worksheet

Instructions: For each observer/repetition, enter into the yellow boxes
the number of times each card appeared in the "less dense" pile.

card number	observers/repetitions												frequency less dense	a) frequency more dense	b) probability more dense	c) z-score more dense
	1	2	3	4	5	6	7	8	9	10	11	12				
1	3	3	3	3	2	3	3	3	3	3	3	3	35	1	0.03	-1.91
2	3	3	3	3	3	3	3	3	2	3	3	3	35	1	0.03	-1.91
3	3	3	2	3	2	3	3	3	3	3	3	3	34	2	0.06	-1.59
4	3	3	3	3	3	3	3	2	3	3	3	2	34	2	0.06	-1.59
5	2	2	1	1		3	2	2	1	2	3	3	22	14	0.39	-0.28
6	2	2				2	2	1		1	2	2	14	22	0.61	0.28
7	3	2				1	2	1	1	1	1	2	14	22	0.61	0.28
8	1						1					2	4	32	0.89	1.22
9													0	36	1.00	#N/A
10							1						1	35	0.97	1.91
11													0	36	1.00	#N/A
12													0	36	1.00	#N/A
13													0	36	1.00	#N/A

(a) frequency "more dense"

(b) probability "more dense"

(c) z-score "more dense"

$y = 0.47x - 2.82$
$R^2 = 0.95$

(d) modeled psychometric function
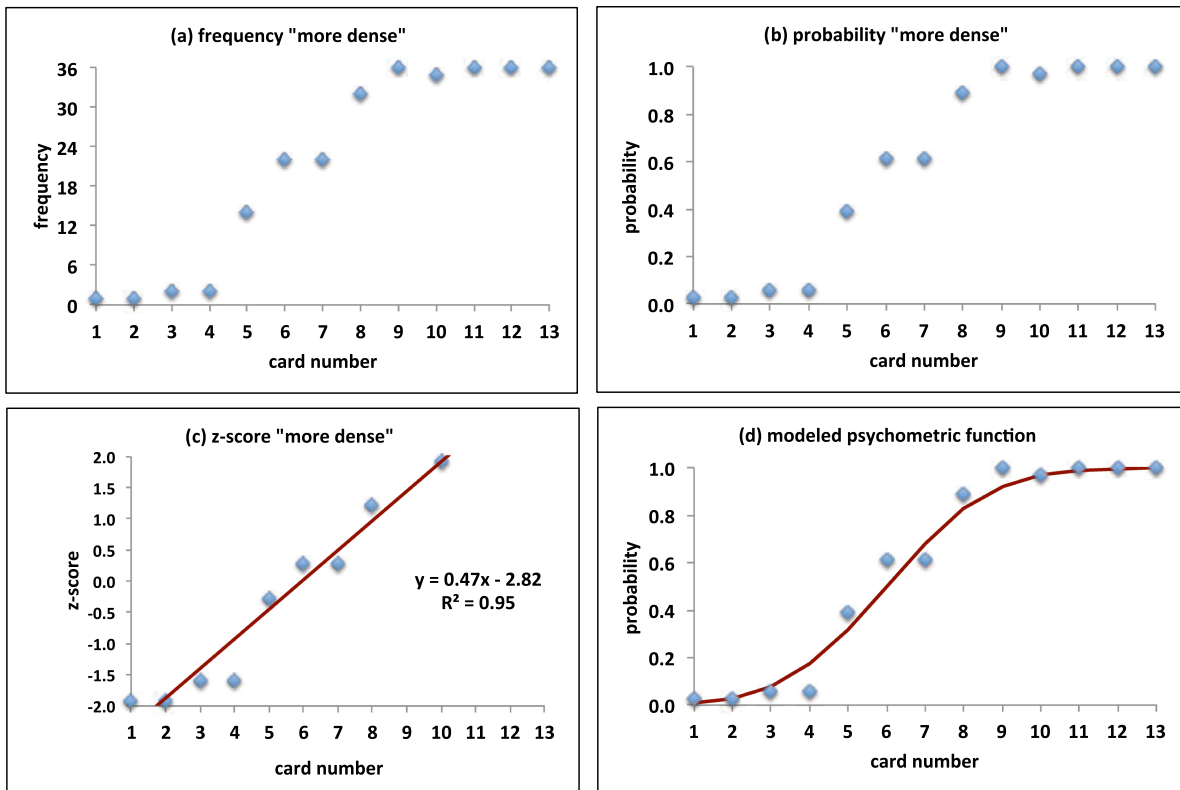
point of subjective equality (PSE) = p(0.50) = 6.00
upper threshold (UL) = p(0.75) = 7.44
lower threshold (LL) = p(0.25) = 4.56
interval of uncertainty (IU) = UL - LL = p(0.75) - p(0.25) = 2.87
upper just-noticeable-difference (JNDu) = UL - PSE = p(0.75) - p(0.50) = 1.44
lower just-noticeable-difference (JNDl) = PSE -LL = p(0.50) - p(0.25) = 1.44
average JND = (JNDu + JNDl)/2 = 1.44

Fig. 4. MCS analysis.

linear relationship to the noise densities. To test this proposition, the spreadsheet plots the z-scores versus the noise densities and then fits a linear function to these data (Figure 4(c)). The appropriateness of a cumulative normal model can be assessed from the coefficient of determination (R^2) of the fit. Proceeding, the mean and standard deviation of the best-fitting cumulative normal model are then calculated from the slope and intercept of the linear function, and the psychometric function model is plotted against the empirical psychometric function (Figure 4(d)).

Finally, the spreadsheet computes a variety of psychophysical measures from the parameters of the model. The PSE is the noise density that corresponds to the $p = 0.5$ point on the model curve. Following convention, the upper and lower discrimination thresholds (UL/LL) are calculated from the $p = 0.75$ and $p = 0.25$ points on the model curve. Finally, the JND is calculated by taking the difference between the PSE and the upper threshold. Given that the cumulative normal model is symmetric around the PSE, the JND could also be estimated by taking the difference with the lower threshold; however, these model-based estimates should be compared to the empirical data to determine their validity.

4.2 Example 2: Paired Comparison

Paired comparison is an experimental method for deriving an interval scale of perceived differences from ordinal judgments of stimulus pairs. On each trial, a pair of stimuli is randomly selected from a discrete test set and presented to the observer. The observer's task is to judge which of the stimuli has more of some attribute than the other. Over repeated presentations, the frequencies with which each of the test stimuli is judged greater than each of the others are tabulated in a matrix for subsequent analysis.

Analysis of the data from a paired comparison experiment is typically done using Thurstone's law of comparative judgment, which proposes that ordinal judgments of stimulus pairs can be described by a discriminal process that is normally distributed. In the simplest case, the discriminal processes associated with different stimuli are assumed to be uniform and uncorrelated. To derive scale values for the stimuli, the values in the frequency (F) matrix created during the experiment are first normalized by the number of times each of the stimulus pairs was presented. This generates a probability (P) matrix with values ranging from 0 to 1. Next, the values in the P matrix are converted to z-scores, producing a Z matrix where each entry can be interpreted as distance from the zero-valued mean of a standard normal distribution. Scale values for each of the stimuli are then found by averaging the z-scores associated with a given stimulus. This procedure produces an interval scale, where the scale values represent the relative perceived magnitudes of the stimuli and differences between scale values indicate the relative discriminability of the stimuli. Although this analysis produces an interval scale with these properties, it must be cautioned that determining meaningful JNDs in scale values requires further analysis [19].

The procedure for running a paired comparison experiment using the FechDeck is illustrated in Figure 5. The instructions are as follows:

1. Sort the deck into suits. Set aside the diamonds.
2. Extract the cards ace through 9 from the spades, clubs, and hearts suits.
3. Shuffle together the spades and clubs to create the "test" pile. Separately shuffle the hearts to create the "standards" pile. Place the two card piles face down on the table.
4. Compare the patterns on each card in the test pile to the top card in the standards pile, sorting the test cards into "less dense" and "more dense" sub-piles.
5. Turn over the top card in the standards pile.
6. Turn over more dense sub-pile. Sort the cards by rank.
7. Enter the number of cards of each rank into the row of the "pair_comparison" spreadsheet that corresponds to the top card in the standards pile.
8. Place the top standards card face up on the bottom of the standards pile. Reassemble, flip, and shuffle the test pile.
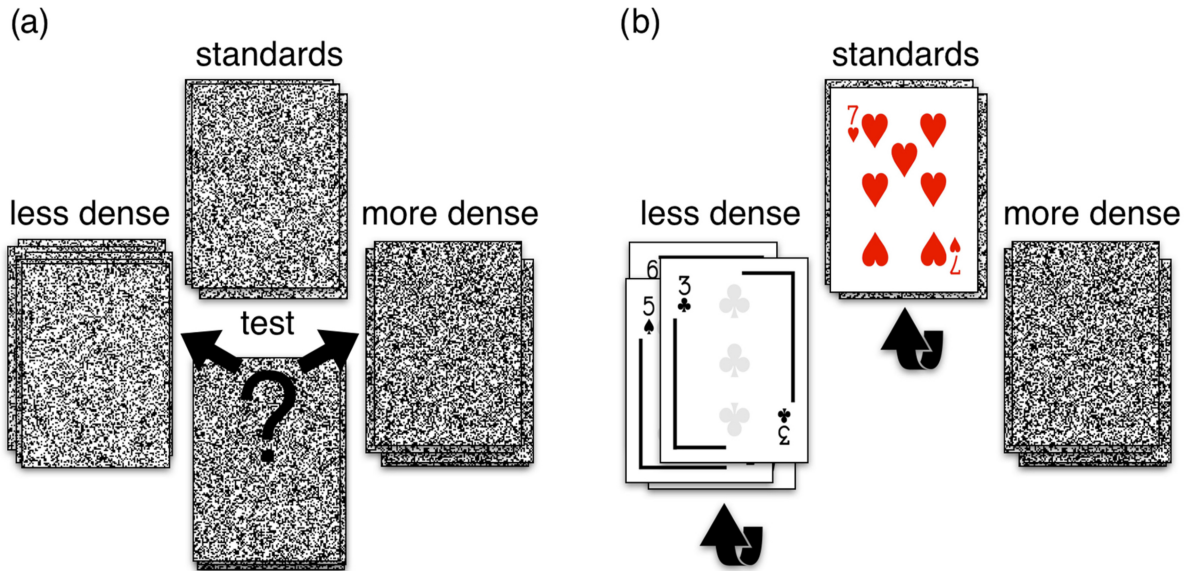
Fig. 5. Paired comparison procedure. (a) Testing. (b) Data gathering.

9. Repeat steps 4 through 8 for each card in the standards pile.
10. Observations/repetitions: (Group) Accumulate each observer's results into the cells in the spreadsheet; (Individual) Repeat steps 3 through 9 until the desired number of repetitions has been completed.
11. Analyze the data using the spreadsheet.

An Excel spreadsheet that performs the data analysis is provided in the FechDeck distribution. The features of this spreadsheet are illustrated in Figure 6. The data show the results from 12 graduate student observers using the FechDeck in a classroom setting.

To perform the analysis, the user enters the "more dense" responses for each test card into the yellow cells in the "raw data" matrix (Figure 6(a)), accumulating values over repetitions. The empty cells in this matrix implicitly represent the test cards that were judged "less dense" than the corresponding standard. The explicit values of these cells can be calculated by subtracting the entries in the "more dense" cells from the total number of presentations. The spreadsheet performs these calculations to generate the complete frequency (F) matrix (Figure 6(b)). The spreadsheet divides the entries in the F matrix by the number of presentations to generate the probability (P) matrix (Figure 6(c)) whose values range from 0 to 1. Next, the entries in the P matrix are converted to z-scores to generate the Z matrix (Figure 6(d)). According to the law of comparative judgment, the entries in the columns of the Z matrix represent the normalized values of the discriminal dispersions of each test card relative to the standards, and the averages of these values yield an interval scale that represents the relative perceived densities of the stimuli. The spreadsheet calculates these column averages and plots the resulting scale values.

4.3 Example 3: Method of Adjustment

The MOA is one of Fechner's "classical" threshold methods. On each trial in an MOA experiment, the observer's task is to adjust a continuously variable test stimulus until it matches a standard. Due to observer variability, the final match value will usually vary from trial to trial.

Analysis of the data, similar to the MCS, can involve fitting a psychometric function, but in practice, analysis of MOA data is usually more basic and consists of averaging the differences between the matched values and

FechDeck pair comparison worksheet

Instructions: For each standard card, enter into the yellow boxes the number of times each test card appears in the "more dense" pile. Accumulate values across observers/repetitions.

(a) raw data		test card								
		1	2	3	4	5	6	7	8	9
standard card	1	4	14	18	23	23	24	23	24	24
	2		12	12	17	22	22	23	24	23
	3			5	17	20	23	22	24	24
	4				13	17	20	23	24	24
	5					16	19	19	22	23
	6						14	22	23	24
	7							16	20	23
	8								15	22
	9									17

(b) F matrix		test card								
		1	2	3	4	5	6	7	8	9
standard card	1	4	14	18	23	23	24	23	24	24
	2	10	12	12	17	22	22	23	24	23
	3	6	12	5	17	20	23	22	24	24
	4	1	7	7	13	17	20	23	24	24
	5	1	2	4	7	16	19	19	22	23
	6	0	2	1	4	5	14	22	23	24
	7	1	1	2	1	5	2	16	20	23
	8	0	0	0	0	2	1	4	15	22
	9	0	1	0	0	1	0	1	2	17

(c) P matrix		test card								
		1	2	3	4	5	6	7	8	9
standard card	1	0.1667	0.5833	0.75	0.9583	0.9583	1	0.9583	1	1
	2	0.4167	0.5	0.5	0.7083	0.9167	0.9167	0.9583	1	0.9583
	3	0.25	0.5	0.2083	0.7083	0.8333	0.9583	0.9167	1	1
	4	0.0417	0.2917	0.2917	0.5417	0.7083	0.8333	0.9583	1	1
	5	0.0417	0.0833	0.1667	0.2917	0.6667	0.7917	0.7917	0.9167	0.9583
	6	0	0.0833	0.0417	0.1667	0.2083	0.5833	0.9167	0.9583	1
	7	0.0417	0.0417	0.0833	0.0417	0.2083	0.0833	0.6667	0.8333	0.9583
	8	0	0	0	0	0.0833	0.0417	0.1667	0.625	0.9167
	9	0	0.0417	0	0	0.0417	0	0.0417	0.0833	0.7083

(d) Z matrix		test card								
		1	2	3	4	5	6	7	8	9
standard card	1	-0.97	0.21	0.67	1.73	1.73		1.73		
	2	-0.21	0.00	0.00	0.55	1.38	1.38	1.73		1.73
	3	-0.67	0.00	-0.81	0.55	0.97	1.73	1.38		
	4	-1.73	-0.55	-0.55	0.10	0.55	0.97	1.73		
	5	-1.73	-1.38	-0.97	-0.55	0.43	0.81	0.81	1.38	1.73
	6		-1.38	-1.73	-0.97	-0.81	0.21	1.38	1.73	
	7	-1.73	-1.73	-1.38	-1.73	-0.81	-1.38	0.43	0.97	1.73
	8					-1.38	-1.73	-0.97	0.32	1.38
	9		-1.73			-1.73		-1.73	-1.38	0.55

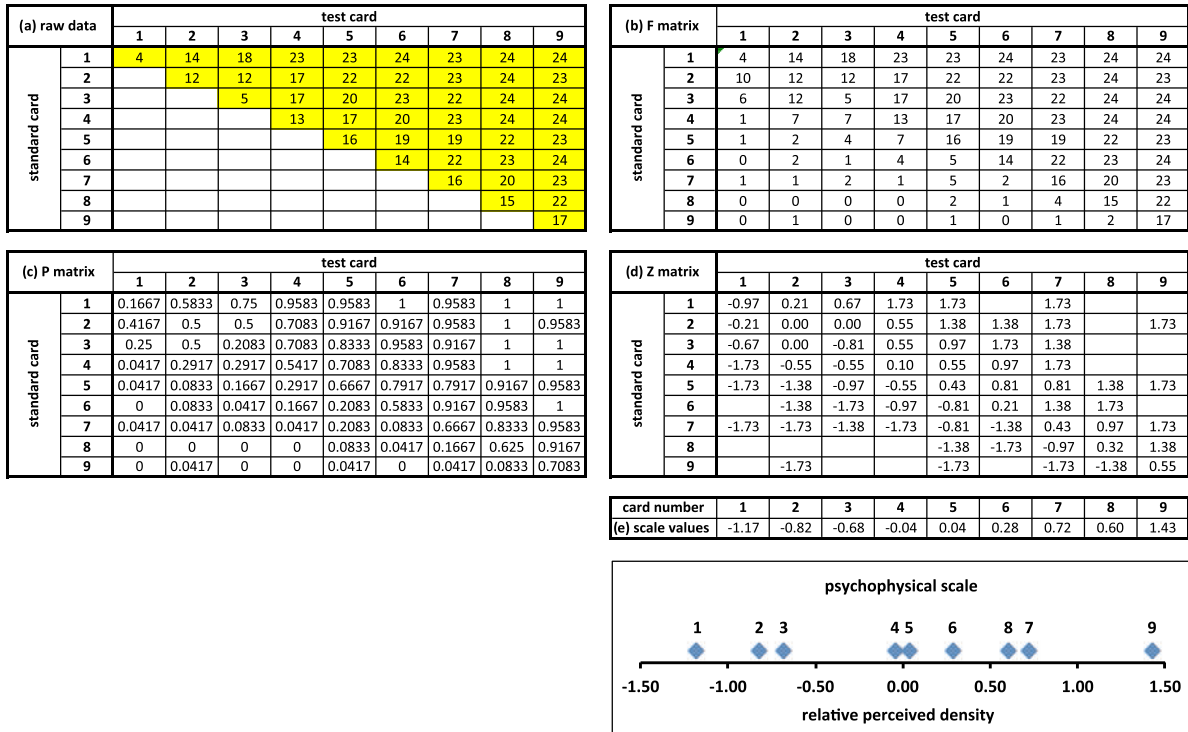| card number	1	2	3	4	5	6	7	8	9
(e) scale values	-1.17	-0.82	-0.68	-0.04	0.04	0.28	0.72	0.60	1.43

Fig. 6. Paired comparison analysis.

the standard value. This average can be used to estimate psychophysical measures such as the PSE and JND. In addition, if a range of standards is used in an experiment, the validity of Weber's law can be tested and a Weber fraction estimated.

Implementing the MOA presented a unique challenge in designing the FechDeck, because the method requires a continuously variable test stimulus, and this seemed impossible given the discrete nature of a card deck. The solution was to develop an experimental task in which the observer's response is indicated by *arranging* the cards rather than *sorting* them.

The procedure for running a MOA experiment using the FechDeck is shown in Figure 7. The instructions are as follows:

1. Extract all cards numbered 2 through 6 from the deck
2. Sort these cards into suits, shuffle each suit pile, and place each pile face down on the table.
3. Pick up the top card in the first suit pile, turn it face up, and place one of the jokers on top of this card face down.
4. Holding the cards horizontally, slide the joker back and forth until the legs of the L-shaped figure in the upper left corner appear to be the same length.
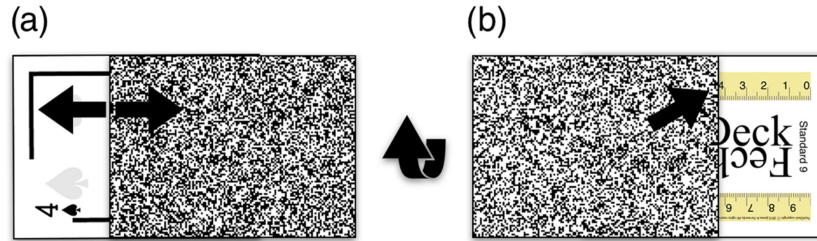5. Hold the cards in place and turn them over.

Fig. 7. MOA procedure. (a) Testing. (b) Data gathering.

FechDeck method of adjustment worksheet

Instructions: For each standard length, enter the matched length into the yellow boxes.

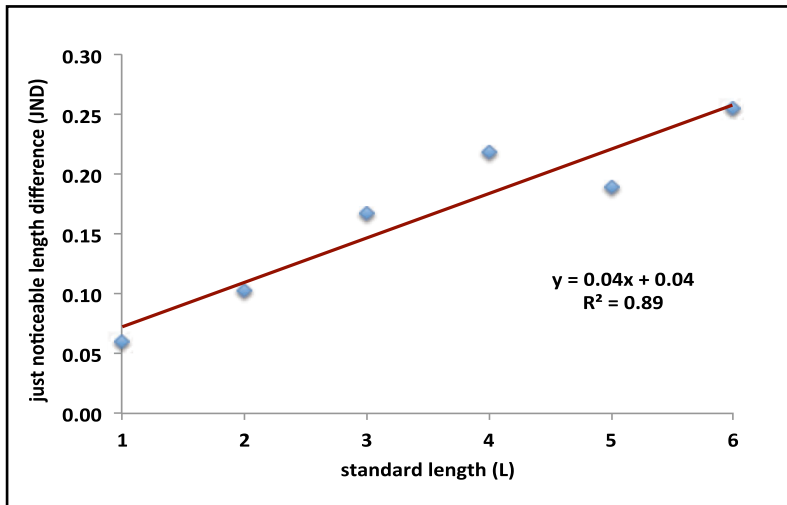| standard length (L)	matched lengths observers/repetitions												mean matched length	JND matched length
	1	2	3	4	5	6	7	8	9	10	11	12		
1	1.0	0.9	0.8	1.1	1.1	1.0	1.0	0.9	1.0	1.0	1.0	1.0	0.979	0.060
2	2.0	1.7	1.9	2.1	2.2	2.0	2.1	1.7	2.1	2.0	2.0	2.0	1.983	0.102
3	2.9	2.9	2.8	3.1	3.6	3.2	3.4	2.8	3.2	3.0	3.0	2.9	3.063	0.167
4	3.8	4.0	3.5	4.3	3.9	4.0	4.5	4.1	4.1	4.1	4.1	3.3	3.975	0.218
5	5.1	5.0	5.5	5.2	5.3	5.2	5.5	5.3	5.3	5.4	4.9	4.5	5.183	0.189
6	5.7	5.6	6.0	6.3	6.0	6.1	6.5	5.5	6.5	6.5	6.0	5.5	6.017	0.255

Fig. 8. MOA analysis.

6. Note the length of the horizontal "match" leg from the top ruler on the face of the joker and the length of the vertical "standard" leg from the number on the face of the other card.
7. Enter the length of the "match" leg into the corresponding "standard length" cell in the "MOA" spreadsheet.
8. Observations/repetitions: (Group) Enter each observer's results in separate columns in the spreadsheet; (Individual) Repeat steps 3 through 7 for the remaining suit piles.
9. Analyze the data using the spreadsheet.

An Excel spreadsheet that performs the data analysis for the MOA experiment is provided in the FechDeck distribution. The features of the spreadsheet are illustrated in Figure 8. The data show the results from 12 graduate student observers using the FechDeck in a classroom setting. To perform the analysis, the user first enters

the "matched length" values for each of the standard lengths (L) into the yellow cells in the spreadsheet. The spreadsheet then calculates the "mean matched length" for each of the standards. This value is an estimate of the PSE of the match line to the corresponding standard. Since the experiment implements a variation on the horizontal-vertical illusion, with sufficient data the users could investigate whether there is any systematic bias in their length estimates.

The spreadsheet then calculates the standard deviations of the matched lengths for each standard and uses the assumptions of the cumulative normal model (illustrated in Example 1: Method of Constant Stimuli) to estimate the JNDs in matched length for each standard. Notably, the magnitudes of the JNDs increase systematically with the length of the standard.

To investigate this relationship, the spreadsheet then visualizes the data, plotting the JNDs for each standard. If the length matching task follows Weber's law, the respective length JND will be a constant fraction of the standard length. To test for this, the spreadsheet fits a linear function to the plotted data. In this case, the goodness of fit of the linear function provides support for Weber's law, and the slope of the line indicates that over the range tested, the Weber fraction is approximately 0.04 (4%).

4.4 Additional Methods

The preceding sections only illustrate a subset of the psychophysical methods supported by the FechDeck, which also includes the methods of limits for threshold measurement, as well as the methods of ranking, successive categories, and magnitude estimation for scaling. Instructions for running these kinds of experiments, and spreadsheets for performing the data analyses, are available through the resources described in the following section.

5 DISSEMINATION

The author's main goal in developing the FechDeck is to provide a set of tools for furthering education in the methods of psychophysics. As such, the components of the FechDeck have been made freely available on the Internet. At the time of writing, there is a GitHub repository at https://github.com/jamesferwerda/FechDeck that contains the card images, as well as the current versions of the instructions and analysis spreadsheets. At the time of this writing, professionally printed versions of the FechDeck are also available from the author at the cost of production.

6 SUMMARY, LIMITATIONS, AND FUTURE WORK

This article has described the design, usage, and dissemination of the FechDeck, a tool to support exploration of psychophysical methods. The FechDeck is an ordinary deck of playing cards that has been modified so that the cards can be used as stimuli in a variety of psychophysical experiments. By manipulating the cards according to instructions provided with the deck, a user can experience the particulars of different threshold and scaling methods and gather data that can be analyzed using the provided spreadsheets to derive standard psychophysical measures.

Although the current version of the FechDeck has been used by the author and colleagues to teach psychophysics in a range of classroom and individualized settings, and has proved to be a popular and useful tool, the current version of the deck has some limitations that should be acknowledged and opportunities for improvement that should be outlined.

First, as in all of psychophysics, there is a constant tension in using the FechDeck between getting enough data to perform meaningful statistical analyses, and not overburdening observers with endless trials. The procedures outlined in the instructions for the experiments were designed primarily to give students an enjoyable, hands-on experience with the methods rather than to gather sufficient data for robust analyses. Therefore, on occasion, some of the assumptions used in the analyses (e.g., normality) may be unfounded. Instructors using

the FechDeck need to make students aware of these limitations and educate students about proper procedures for data gathering and statistical analysis.

In addition, although the current version of the FechDeck supports many of the "classical" and best-known psychophysical methods, future versions of the deck could be expanded by developing procedures and analysis tools that support a wider range of methods. Given the generic nature of the FechDeck noise pattern stimuli, this seems entirely possible, at least for some classes of methods. For example, with a little more thinking, it seems like it should be possible to develop instructions to support "staircase" and other adaptive threshold methods [7], as well as modern scaling techniques such as maximum likelihood difference scaling [20]. In addition, analysis spreadsheets that demonstrate more modern approaches to choosing and fitting psychometric functions such as those described in other works [8–10, 18, 21, 22] would be valuable additions for more advanced students. Support for experiments that explore the theory of signal detection [23] would also be an important priority, and given that signal detection focuses on the effects of externalities on experimental findings, it should be possible to support the teaching of these concepts by developing instructions that vary the experimental conditions (e.g., stimulus frequencies) of currently supported methods. Finally, there are some important modern psychophysical methods such as QUEST [24, 25] and its peers that seem harder to support with the FechDeck due to the continuous nature of adaptive stimulus selection; however, as with support for the MOA in the current deck, motivated users in the open source community may find clever ways to extend the deck to support these and other methods.

ACKNOWLEDGMENTS

REFERENCES

[1] G. A. Gescheider. 1997. *Psychophysics: The Fundamentals*. Taylor & Francis Group.

[2] Z. L. Lu and B. Dosher. 2013. *Visual Psychophysics: From Laboratory to Theory*. MIT Press, Cambridge, MA.

[3] F. A. A. Kingdom and N. Prins. 2016. *Psychophysics: A Practical Introduction*. Academic Press.

[4] G. T. Fechner. 1860. *Elemente der Psychophysik*. Breitkopf und Härtel. Translated/published 1966 as *Elements of Psychophysics*. Holt, Rinehart, & Winston.

[5] L. L. Thurstone. 1927. A law of comparative judgment. *Psychological Review* 34, 4 (1927), 273–286.

[6] S. S. Stevens. 1957. On the psychophysical law. *Psychological Review* 64, 3 (1957), 153–181.

[7] B. Treutwein. 1995. Adaptive psychophysical procedures. *Vision Research* 35, 17 (1995), 2503–2522.

[8] B. Treutwein and H. Strasburger. 1999. Fitting the psychometric function. *Perception and Psychophysics* 61, 1 (1999), 87–106.

[9] S. A. Klein. 2001. Measuring, estimating, and understanding the psychometric function: A commentary. *Perception and Psychophysics* 63, 8 (2001), 1421–1455.

[10] N. A. Macmillan. 2001. Threshold estimation: The state of the art. *Perception and Psychophysics* 63, 8 (2001), 1277–1278.

[11] F. M. Urban. 1910. The method of constant stimuli and its generalizations. *Psychological Review* 17, 4 (1910), 229–259.

[12] D. H. Brainard. 1997. The Psychophysics Toolbox. *Spatial Vision* 10, 4 (1997), 433–436.

[13] M. Kleiner, D. Brainard, D. Pelli, A. Ingling, and R. Murray. 2007. What's new in Psychtoolbox-3. *Perception* 36, 14 (2007), 1–16.

[14] J. W. Peirce. 2007. PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods* 162 (2007), 8–13.

[15] A. D. Straw. 2008. Vision Egg: An open-source library for real-time visual stimulus generation. *Frontiers in Neuroinformatics* 2 (2008), 1–10.

[16] H. Strasburger. 2018. Software for visual psychophysics: An overview. V 9.4. Available at http://www.visionscience.com.

[17] MathWorks. 2017. Image Processing Toolbox. Available at https://www.mathworks.com.

[18] F. A. Wichmann and N. J. Hill. 2001. The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception and Psychophysics* 63, 8 (2001), 1293–1313.

[19] A. B. Watson and L. Kreslake. 2001. Measurement of visual impairment scales for digital video. In *Proceedings of SPIE Human Vision and Electronic Imaging VI*, Vol. 4299. 79–89.

[20] L. T. Maloney and J. N. Yang. 2003. Maximum likelihood difference scaling. *Journal of Vision* 3, 8 (2003), 5.

[21] H. Strasburger. 2001. Converting between measures of slope of the psychometric function. *Perception and Psychophysics* 63, 8 (2001), 1348–1355.

[22] H. H Schütt, S. Harmeling, J. H. Macke, F. A. Wichmann. 2016. Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Research* 122 (2016), 105–123.

[23] D. M. Green and J. A. Swets. 1966. *Signal Detection Theory and Psychophysics*. Wiley.

[24] A. B. Watson and D. G. Pelli. 1983. Quest: A Bayesian adaptive psychometric method. *Perception and Psychophysics* 33, 2 (1983), 113–120.

[25] A. B. Watson. 2017. QUEST+: A general multidimensional Bayesian adaptive psychometric method. *Journal of Vision* 17, 3 (2017), 10.