Title
Evaluating the performance of the staircase and quick Change Detection methods in measuring perceptual learning

Journal
Journal of Vision, 19(7)

ISSN
1534-7362

Authors
Zhang, Pan
Zhao, Yukai
Dosher, Barbara Anne
et al.

Publication Date
2019-07-19

DOI
10.1167/19.7.14

Evaluating the performance of the staircase and quick Change Detection methods in measuring perceptual learning

Pan Zhang
Laboratory of Brain Processes (LOBES), Department of Psychology, The Ohio State University, Columbus, OH, USA

Yukai Zhao
Laboratory of Brain Processes (LOBES), Department of Psychology, The Ohio State University, Columbus, OH, USA

Barbara Anne Dosher
Department of Cognitive Sciences and Institute of Mathematical Behavioral Sciences, University of California, Irvine, CA, USA

Zhong-Lin Lu
Laboratory of Brain Processes (LOBES), Department of Psychology, The Ohio State University, Columbus, OH, USA

The staircase method has been widely used in measuring perceptual learning. Recently, Zhao, Lesmes, and Lu (2017, 2019) developed the quick Change Detection (qCD) method and applied it to measure the trial-by-trial time course of dark adaptation. In the current study, we conducted two simulations to evaluate the performance of the 3-down/1-up staircase and qCD methods in measuring perceptual learning in a two-alternative forced-choice task. In Study 1, three observers with different time constants (40, 80, and 160 trials) of an exponential learning curve were simulated. Each simulated observer completed staircases with six step sizes (1%, 5%, 10%, 20%, 30%, and 60%) and a qCD procedure, each starting at five levels (+50%, +25%, 0, −25%, and −50% different from the true threshold in the first trial). We found the following results: Staircases with 1% and 5% step sizes failed to generate more than five reversals half of the time; and the bias and standard deviations of thresholds estimated from the post hoc segment-by-segment qCD analysis were much smaller than those from the staircase method with the other four step sizes. In Study 2, we simulated thresholds in the transfer phases with the same time constants and 50% transfer for each observer in Study 1. We found that the estimated transfer indexes from qCD showed smaller biases and standard deviations than those from the staircase method. In addition, rescoring the simulated data from the staircase method using the Bayesian estimation component of the qCD method resulted in much-improved estimates. We conclude that the qCD method characterizes the time course of perceptual learning and transfer more accurately, precisely, and efficiently than the staircase method, even with the optimal 10% step size.

Introduction

Perceptual learning refers to the improvement in an observer's performance in a perceptual task through training or practice. It has been documented in all sensory modalities (vision: Goldstone, 1998; Lu, Hua, Huang, Zhou, & Dosher, 2011; Sagi, 2011; Sasaki, Náñez, & Watanabe, 2010; Sasaki, Náñez, & Watanabe, 2012; Watanabe & Sasaki, 2015; hearing: Banai & Amitay, 2012; Moore, Amitay, & Hawkey, 2003; Wright & Zhang, 2009; smell: Stevenson, 2001; Wilson & Stevenson, 2003; taste: Blair & Hall, 2003; Mackintosh, Kaye, & Bennett, 1991; Scahill & Mackintosh, 2004; Symonds & Hall, 1995; touch: Rodríguez & Angulo, 2014; Sathian & Zangaladze, 1998). Perceptual learning not only unlocks important plasticity of the perceptual system (Petrov, Dosher, & Lu, 2005; Sagi, 2011; Sasaki et al., 2012) but also provides noninvasive rehabilitation methods for a variety of perceptual impairments such as amblyopia (C.-B. Huang, Zhou, & Lu, 2008; Polat, Ma-Naim, Belkin, & Sagi, 2004), myopia (Camilleri, Pavan, Ghin, & Campana, 2014;

Citation: Zhang, P., Zhao, Y., Dosher, B. A., & Lu, Z.-L. (2019). Evaluating the performance of the staircase and quick Change Detection methods in measuring perceptual learning. *Journal of Vision*, *19*(7):14, 1–25, https://doi.org/10.1167/19.7.14.

https://doi.org/10.1167/19.7.14 Received January 25, 2019; published July 19, 2019 ISSN 1534-7362

Tan & Fong, 2008; Yan et al., 2015), hemianopia (Nelles et al., 2001; Perez & Chokron, 2014), and aging (Andersen, Ni, Bower, & Watanabe, 2010; Bower & Andersen, 2012; Bower, Watanabe, & Andersen, 2013).

One fundamental empirical phenomenon in perceptual learning is the learning curve. Most of the important discoveries about perceptual learning are based on the analysis of the learning curve in the training phase or the transfer phase. Indeed, the very existence of perceptual learning is based on whether the slope of the learning curve is significantly different from zero (Dosher & Lu, 2007; C.-B. Huang et al., 2008; Polat et al., 2004; Yan et al., 2015; Zhang, Hou, et al., 2018; Zhou et al., 2006). Changes of the learning curve obtained under different experimental manipulations such as external noise (Lu, Chu, & Dosher, 2006; Lu & Dosher, 2004), training difficulty (J. Liu, Lu, & Dosher, 2010, 2012; Z. Liu & Weinshall, 2000), training schedule (Hung & Seitz, 2014; Xiao et al., 2008), feedback (Aberg & Herzog, 2012; Fahle & Edelman, 1993; J. Liu et al., 2010, 2012; Shibata, Yamagishi, Ishii, & Kawato, 2009), attention (Donovan & Carrasco, 2015; Mukai et al., 2007; Szpiro & Carrasco, 2015), and reward (Seitz, Kim, & Watanabe, 2009; Zhang, Hou, et al., 2018) have revealed important properties of perceptual learning and are the bases for the development of theories and computational models of perceptual learning (Dosher & Lu, 2017). One of the hallmark properties of perceptual learning—specificity—is also derived from the learning curves during training and transfer (Ahissar & Hochstein, 1997; Fahle & Morgan, 1996; Z. Liu & Weinshall, 2000; Xiao et al., 2008).

The learning curve in perceptual-learning studies is typically measured with the method of constant stimuli that estimates percentage correct (Fahle, Edelman, & Poggio, 1995; Fahle & Morgan, 1996), with d' (Ball & Sekuler, 1982, 1987), or with adaptive procedures that estimate either contrast thresholds or difference thresholds in blocks of trials (T. Bi, Chen, Zhou, He, & Fang, 2014; Donovan & Carrasco, 2015; Liang, Zhou, Fahle, & Liu, 2015; Polat et al., 2004; Wang et al., 2016; Xiao et al., 2008) with various forms of the staircase procedure (Cornsweet, 1962; Watson & Pelli, 1983) as the most frequently used method (Leek, 2001; Meese, 1995; Monsen & Horn, 2007).

The up-down staircase method is designed to estimate the threshold at a fixed performance level in a block of trials. Before each measurement, the initial stimulus level and the initial step size of the staircase are determined by the experimenter. The method adjusts the stimulus level and the step size based on the observer's performance and some predefined rules. For example, in a 3-down/1-up staircase procedure, the stimulus level is increased by one step if the observer makes one mistake (1 up) and is decreased by one step if the observer makes three consecutive correct responses (3 down). A reversal results when the direction of stimulus change is reversed (from up to down or down to up). In addition, the step size may decrease with the number of reversals. Typically, the staircase procedure requires a certain number of trials to generate enough reversals to provide a relatively precise estimate of the threshold. The performance of the staircase depends on the initial stimulus level and step size (Lu & Dosher, 2013).

One major assumption underlying the staircase procedure is that the threshold does not change during the measurement block, which typically contains about 60–80 trials (Harris et al., 2015; Schoups, Vogels, & Orban, 1995; Sowden, Rose, & Davies, 2002; Thurston & Dobkins, 2007; Xiao et al., 2008). The staircase procedure has been widely used to measure time-varying phenomena such as adaptation (Bao, Fast, Mesik, & Engel, 2013; Binns, Taylor, Edwards, & Crabb, 2018; Jackson, Owsley, & McGwin, 1999; Ward, Morison, Simmers, & Shahani, 2018) and perceptual learning (Badiru, 1992; Heathcote, Brown, & Mewhort, 2000; Lu et al., 2011; Mazur & Hastie, 1978; Petrov et al., 2005). However, there has been no systematic study of the performance of the staircase procedure in measuring learning curves in perceptual learning, including the convergence properties of the procedure, its precision and bias, and the effects of the initial stimulus level and step size, in relation to different rates of learning. One goal of the current study is to systematically evaluate the performance of the staircase procedure in measuring perceptual learning and its transfer.

Another goal of the current study is to compare the performance of the staircase procedure to that of a newly developed Bayesian adaptive procedure, the quick Change Detection (qCD) method, that has been shown to provide accurate and precise assessment of the dark-adaption curve (Zhao, Lesmes, & Lu, 2017, 2019) in an eight-alternative forced-choice (8AFC) task and perceptual learning in a four-alternative forced-choice (4AFC) task (Zhang, Zhao, Dosher, & Lu, 2018, 2019). Here we evaluate the performance of the qCD method in measuring the learning curves during both the learning and transfer phases of a two-alternative forced-choice (2AFC) experiment and compare it directly with that of the staircase procedure. This is quite important for two reasons: 2AFC is the most widely used procedure in perceptual-learning studies (a Google Scholar search on "perceptual learning and staircase" resulted in 3,380 links; Jogan & Stocker, 2014; Kingdom & Prins, 2010; Vancleef et al., 2018), and the number of alternatives in an mAFC procedure strongly affects the performance of the adaptive procedures (J. Bi, Lee, & O'Mahony, 2010; Hall, 1983; Hou, Lesmes, Bex, Dorr, & Lu, 2015; Shelton & Scarrow, 1984). Forced-choice procedures with three or four alternatives provide more satisfactory measure-

ment of psychometric performance (Leek, 2001). Hou et al. (2015) found that increasing the number of alternatives in a forced-choice task greatly improved the efficiency of assessing a contrast sensitivity function in both simulation and psychophysical studies. Previously, we have used 4AFC and 8AFC tasks in comparing the staircase and qCD methods (Zhang, Zhao, et al., 2018, 2019; Zhao et al., 2017, 2019).

The current article consists of two simulation studies. In Simulation Study 1, we evaluated the performance of the 3-down/1-up staircase and qCD methods in measuring the learning curves of three simulated observers with different learning rates. Five initial stimulus levels (+50%, +25%, 0, −25%, and −50% different from the true initial threshold) and six initial step sizes (1%, 5%, 10%, 20%, 30%, and 60% change) were included in the staircase procedure. Uniform priors on the parameters of the exponential curve were used in the qCD method. In Simulation Study 2, we conducted additional simulations to evaluate the accuracy and precision of the estimated transfer index from the qCD and staircase methods.

Simulation Study 1

Methods

To evaluate the performance of the staircase and qCD methods in measuring the perceptual learning curve, we simulated three observers with exponential learning curves at three different rates of learning (Table 1) in a 2AFC task.

The qCD method

In the qCD method (Figure 1; Appendix A), the true learning curve is characterized with an exponential function:

$$T\left(\vec{\theta},n\right)=\lambda\exp\left(\frac{-n}{\gamma}\right)+\alpha, \quad (1)$$

where $\vec{\theta}=(\lambda,\gamma,\alpha)$ are the exponential parameters, n is the trial number during training, and $T(\vec{\theta},n)$ is the threshold in trial n. In a 2AFC task, the probability-correct psychometric function is described by a Weibull function (Wichmann & Hill, 2001):

$$p_n'\left(r=1|\vec{\theta},x\right)= g+(1-g)\left(1-\exp\left(-\left(\frac{x}{T_w\left(\vec{\theta},n\right)}\right)^s\right)\right), \quad (2a)$$

	Observer 1	Observer 2	Observer 3
λ	0.2685	0.2685	0.2685
γ	40	80	160
α	0.0895	0.0895	0.0895

Table 1. Parameters of the three simulated observers in Simulation Study 1.

$$p_n\left(r=1|\vec{\theta},x\right)=(1-l)p_n'\left(r=1|\vec{\theta},x\right)+lg, \quad (2b)$$

where p_n' is the psychometric function on trial n without a lapse rate and p_n is the psychometric function on trial n with a lapse rate, r is the response (1 for correct, 0 for incorrect), x is the stimulus value, and $g=0.5$ is the guessing rate. The slope of the Weibull function was set as $s=3.06$ and the lapse rate was set at $l=0.04$.

A prior distribution $p_0(\vec{\theta})$ (see Appendix A for details) is used to represent a priori knowledge of the parameters of the learning curve at the beginning of the experiment. The qCD method selects the stimulus on each trial from all possible stimulus values to optimize the expected information gain (Kujala & Lukka, 2006; Lesmes, Lu, Baek, & Albright, 2010) and updates the joint posterior distribution based on the response of the observer via Bayes's rule. Stimulus selection and Bayesian updating are iterated until the total number of training trials is completed. Based on the joint posterior distribution after each trial, the qCD provides trial-by-trial estimates of the parameters of the perceptual-learning function and a single threshold estimate. In addition, a post hoc analysis can be performed on the trial-by-trial data to investigate the functional form of the learning curve, and this information can be used to further improve the accuracy and precision of the estimated thresholds by aggregating information across all trials (Zhao et al., 2017, 2019).

To simulate the qCD method, a broad joint prior distribution $p_0(\vec{\theta})$ was defined in the three-dimensional parameter space of $\vec{\theta}=(\theta_1,\theta_2,\theta_3)=(\lambda,\gamma,\alpha)$, with the subscript 0 denoting that the prior represents a priori knowledge of the parameters of the learning curve before the experiment (Appendix A). The parameter space included 50 log-linearly spaced λ values (from 0.1 to 0.4), 50 log-linearly spaced γ values (from 20 to 200), and 50 log-linearly spaced α values (from 0.05 to 0.2). Both λ and α are in the units of the threshold measurement, while λ values are in the unit of trial number. The joint prior distribution $p_0(\vec{\theta})$ was updated trial by trial throughout the simulated experiment.

The 3-down/1-up staircase procedure

The 3-down/1-up staircase method (Cornsweet, 1962) was simulated with five starting stimulus levels (+50%, +25%, 0%, −25%, and −50% from the true

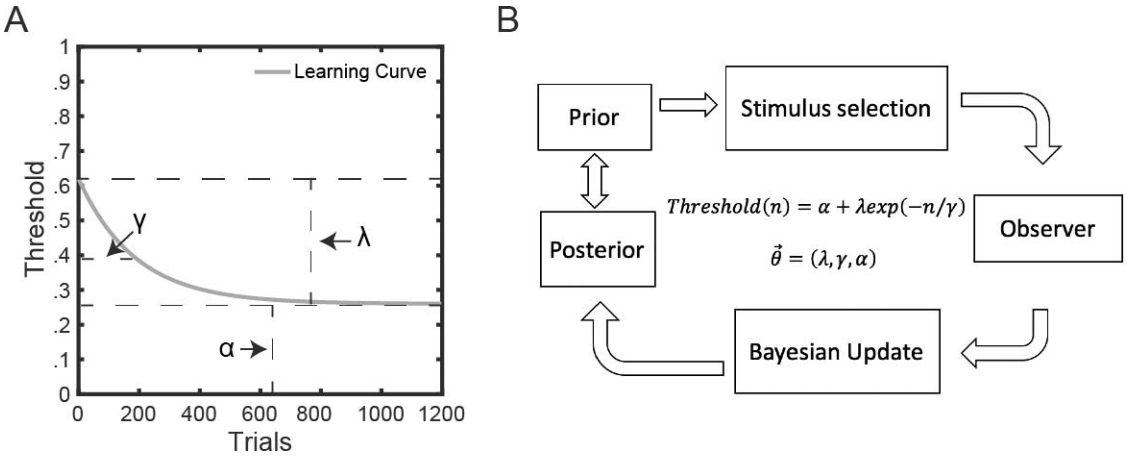

Figure 1. (A) An exponential function with three parameters: λ, the dynamic range of learning; γ, the time constant of the exponential function; and α, the asymptotic performance level. (B) The quick Change Detection method consists of five steps: (1) The learning curve—i.e., the time course of threshold—is defined as an exponential function (as a function of trial number n) with three parameters, $\theta = (\alpha, \gamma, \lambda)$, and their joint prior distribution. (2) The stimulus level in the next trial is selected to optimize the expected information gain on the joint distribution of the parameters. (3) The posterior distribution is updated by Bayes's rule based on observer's response after each trial. (4) Steps 2 and 3 are repeated until the stop criterion is met (e.g., predetermined number of trials). (5) Trial-by-trial and post hoc segment-by-segment thresholds are computed from the posterior distributions.

initial threshold) and six step sizes (1%, 5%, 10%, 20%, 30%, and 60% of the current stimulus level). We denote the staircases with their step sizes—for example, SC10% represents a staircase with a 10% step size. In the 3-down/1-up staircase procedure, if the simulated observer makes three consecutive correct responses, the stimulus level (e.g., contrast, luminance, orientation) is reduced by the step size (multiplied by 1 − step size) in the next trial (3 down). If the simulated observer makes a single incorrect response, the stimulus level is increased by the step size (multiplied by 1 + step size) in the next trial (1 up). A reversal (or endpoint) happens if the staircase changes its direction (from upward to downward or vice versa). Usually, one block of 3-down/1-up staircase with a 10% step size produces about a dozen reversals in 80 trials. The estimated threshold of the block is calculated by averaging stimulus levels at even numbers of reversals after deleting the first four or five of them.

Simulation procedures

We simulated 1,000 runs of 800 trials each for each observer with each of the 35 measurement procedures—qCD and the 3-down/1-up staircase with six step sizes, each starting at one of five starting stimulus levels—resulting in 28,000,000 simulated trials per observer. Each measurement block of the staircase method consisted of 40, 80, or 160 trials. In order to investigate the effects of the initial stimulus level on its performance, we did not use optimal stimulus selection in the first trial of the qCD procedure. Instead, we matched the initial stimulus level in the qCD and staircase procedures.

In each trial, the true threshold $T(n)$ was calculated using Equation 1. The expected probability of a correct response was calculated using Equation 2b. To determine whether the observer's response was correct, we first drew a random number ρ from a uniform distribution over the interval from 0 to 1 and then labeled the response as correct if $\rho < p_n(r = 1|\vec{\theta}, x)$, and incorrect otherwise.

Evaluation metrics

Accuracy is a measure of the difference between the estimated and true values, usually computed as the bias of the estimates. Precision is gauged by the variability of the estimates (Lu & Dosher, 2013; Treutwein, 1995). We computed the bias and variability of the estimated thresholds and parameters of the learning curves. Logarithmic-scale ($\log_{10}$ unit) measures were used in the following analyses because this allowed us to compare the quality of measures in different dimensions.

Accuracy: For the qCD method, the bias (on a $\log_{10}$ scale) of the estimated threshold of the nth simulated trial is defined as

$$Bias_n = \frac{\sum_j \left(log_{10}\left(T_{nj}\right) - log_{10}\left(T_{n,true}\right)\right)}{J}, \quad (3)$$

where T_{nj} is the estimated threshold of the nth trial in the jth simulation and $T_{n,true}$ is the true threshold in the nth trial.

To compute a summary statistic of the bias over the entire learning curve, we cannot simply compute the mean of $Bias_n$, because biases at different time points may carry opposite signs. Instead, we use the root mean square error (RMSE) to represent the average bias over the entire learning curve:

$$RMSE = \sqrt{\frac{\sum_n \left(log_{10}(\bar{T}_n) - log_{10}\left(T_{n,true}\right)\right)^2}{N}}, \quad (4)$$

$$\bar{T}_n = \frac{\sum_j T_{nj}}{J}, \quad (5)$$

where N is the total number of trials in each run.

The staircase method can only provide a single threshold estimate in each measurement block. Bias (on a $\log_{10}$ scale) of the estimated threshold of the bth simulated block is defined as

$$Bias_b = \frac{\sum_j \left(log_{10}\left(T_{bj}\right) - log_{10}\left(T_{b,true}\right)\right)}{J}, \quad (6)$$

where T_{bj} is the estimated threshold in the bth block of the jth simulation, J is the total number of simulations, and $T_{b,true}$ is the true threshold in the middle of the bth block—that is, the threshold in trial $(b - 0.5) \times blocksize$ on the simulated learning curve.

The RMSE of the estimated thresholds of the entire learning curve from the staircase method is defined as

$$RMSE = \sqrt{\frac{\sum_b \left(log_{10}(\bar{T}_b) - log_{10}\left(T_{b,true}\right)\right)^2}{B}}, \quad (7)$$

$$\bar{T}_b = \frac{\sum_j T_{bj}}{J}, \quad (8)$$

where B is the total number of blocks.

The biases of the estimated parameters $\theta_f = (\lambda, \gamma, \alpha)$ from the qCD method are defined as

$$Bias_{\left(n,\theta_f\right)} = \frac{\sum_j \left(log_{10}\left(\theta_{f,nj}\right) - log_{10}\left(\theta_{f,true}\right)\right)}{J}, \quad (9)$$

where $\theta_{f,nj}$ is the estimated parameter value after the nth trial in the jth simulation, J is the total number of simulations, $\theta_{f,true}$ is the true parameter value, and $f = 1$, 2, 3.

For the staircase method, the estimated parameters were obtained from the best-fitting exponential function to the estimated block-by-block thresholds in each run using the MATLAB (MathWorks, Natick, MA) function fminsearch. The biases of the estimated parameters from the staircase method are defined as

$$Bias_{\left(\theta_f\right)} = \frac{\sum_j \left(log_{10}\left(\theta_{f,j}\right) - log_{10}\left(\theta_{f,true}\right)\right)}{J}. \quad (10)$$

Precision: We evaluated the precision of the procedures in two different ways. The first estimate is the standard deviation (SD) of repeated measures. For the qCD method, the SD of the estimated threshold of the nth simulated trial is defined as

$$SD_n = \sqrt{\frac{\sum_j \left(log_{10}\left(T_{nj}\right) - log_{10}(\bar{T}_n)\right)^2}{J}}. \quad (11)$$

The average SD of the estimated threshold of the entire learning curve is defined as

$$\overline{SD}_n = \sqrt{\frac{\sum_n \left(SD_n\right)^2}{N}}. \quad (12)$$

The staircase method can only provide block-by-block threshold estimates. The SD of the estimated thresholds in the bth block is defined as

$$SD_b = \sqrt{\frac{\sum_j \left(log_{10}\left(T_{bj}\right) - log_{10}(\bar{T}_b)\right)^2}{J}}, \quad (13)$$

$$\bar{T}_b = \frac{\sum_j T_{bj}}{J}. \quad (14)$$

The average SD of the estimated threshold of the entire learning curve from the staircase method is defined as

$$\overline{SD}_b = \sqrt{\frac{\sum_b \left(SD_b\right)^2}{B}}, \quad (15)$$

where B is the total number of blocks in each run.

Similarly, the SDs of the estimated parameters from the qCD method are defined as

$$SD_{\left(n,\theta_f\right)} = \sqrt{\frac{\sum_j \left(log_{10}\left(\theta_{f,nj}\right) - log_{10}\left(\bar{\theta}_{f,n}\right)\right)^2}{J}}, \quad (16)$$

$$\bar{\theta}_f = \frac{\sum_j \theta_{fj}}{J}. \quad (17)$$

The SDs of the estimated parameters from the staircase method are defined as

$$SD_{\left(\theta_f\right)} = \sqrt{\frac{\sum_j \left(log_{10}\left(\theta_{f,j}\right) - log_{10}\left(\bar{\theta}_f\right)\right)^2}{J}}. \quad (18)$$

The second way to gauge precision is by the half width of the 68.2% credible interval (HWCI) of the posterior distributions (Edwards, Lindman, & Savage, 1963). The HWCI can be used to assess the precision of a single run of the qCD procedure (Greenland & Kenneth, 1997). For example, the 68.2% credible interval approximates the range within which the actual value lies with 68.2% probability. This is especially

	Observer 1					Observer 2					Observer 3				
Step size	0%	+25%	−25%	+50%	−50%	0%	+25%	−25%	+50%	−50%	0%	+25%	−25%	+50%	−50%
1%	49.6%	22.6%	88.2%	17.7%	99.6%	93.9%	45.2%	99.5%	22.8%	99.5%	97%	87.5%	100%	35.2%	99.8%
5%	98.5%	88.8%	99.4%	50.9%	99.6%	100%	99.9%	100%	99.5%	100%	100%	100%	100%	99.9%	100%
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
30%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
60%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Table 2. Percentage of staircase runs with more than five reversals in the first block of learning.

important for perceptual-learning studies, because the learning curve can only be measured once in an individual; and therefore estimating *SD* from repeated experimental runs is not possible.

Results

We tested the qCD and staircase methods with five starting stimulus levels: +50%, +25%, 0%, −25%, and −50% from the true threshold in the first trial. Because the patterns of results with the five starting levels are quite similar, we mainly present the results with the 0% starting level in the main text. The results with the other four starting levels are presented in the supplementary information.

Staircase convergence

In the 3-down/1-up staircase procedure, the estimated threshold in each measurement block is calculated by averaging stimulus levels at an even number of reversals after deleting the first four or five of them. We computed the block thresholds from the endpoints in every 40, 80, or 160 trials. However, because many staircases with a block size of 40 did not generate more than five reversals (see Supplementary Table S1) and staircases with a block size of 160 produced an insufficient number of block thresholds (see Supplementary Figure S1), we only evaluated the performance of the staircase method with a block size of 80 trials in the following analyses. Table 2 shows the percentage of staircases with more than five reversals in the first measurement block of the learning curves. With step sizes of 1% and 5%, many staircases didn't generate more than five reversals. We therefore dropped these two conditions in subsequent analyses. In summary, we evaluated the performance of the staircase method only with four step sizes (10%, 20%, 30%, and 60%) and one block size (80 trials).

Estimated learning curves

The estimated learning curves in the first 300 trials in the trial-by-trial (red line) and post hoc segment-by-segment (blue line) analyses from the qCD method, and the block-by-block threshold estimates from the staircase method with a 0% starting level, are shown in Figure 2. Visual inspection suggests that the estimated post hoc segment-by-segment thresholds from the qCD method were very close to the true thresholds, and closer than the block-by-block thresholds from the staircase method (SC10%, SC20%, SC30%, and SC60%). In addition, the estimated thresholds from the qCD method were more precise than those from the staircase method. We quantify these observations next. The estimated thresholds with ±25% and ±50% starting levels are shown in Supplementary Figures S2 through S5.

Accuracy and precision of the estimated thresholds

The biases of the estimated thresholds from the qCD and staircase methods with the 0% starting level are plotted in Figure 3A. The RMSEs of the estimated thresholds from the qCD trial-by-trial, qCD post hoc segment-by-segment, SC10%, SC20%, SC30%, and SC60% were 0.015, 0.005, 0.011, 0.014, 0.028, and 0.090 $\log_{10}$ units, respectively, for Observer 1; 0.015, 0.004, 0.007, 0.015, 0.030, and 0.092 $\log_{10}$ units for Observer 2; and 0.018, 0.002, 0.003, 0.015, 0.030, and 0.093 $\log_{10}$ units for Observer 3. The qCD method was much more accurate than the staircase method. The RMSEs of the estimated thresholds with starting levels of ±25% and ±50% are given in Table 3; more details are provided in Supplementary Figure S6.

The *SD*s of the estimated thresholds from the qCD and staircase methods with the 0% starting level are shown in Figure 3B. The *SD*s of the estimated thresholds from the qCD trial-by-trial, qCD post hoc segment-by-segment, SC10%, SC20%, SC30%, and SC60% were 0.029, 0.017, 0.045, 0.052, 0.059, and 0.084 $\log_{10}$ units, respectively, for Observer 1; 0.032, 0.018, 0.044, 0.052, 0.059, and 0.085 $\log_{10}$ units for Observer 2; and 0.032, 0.017, 0.043, 0.052, 0.059, and 0.085 $\log_{10}$ units for Observer 3. For the staircase methods, the *SD* increased with step size. The *SD*s of the estimated thresholds from the qCD method were always considerably smaller than those from the staircase methods.

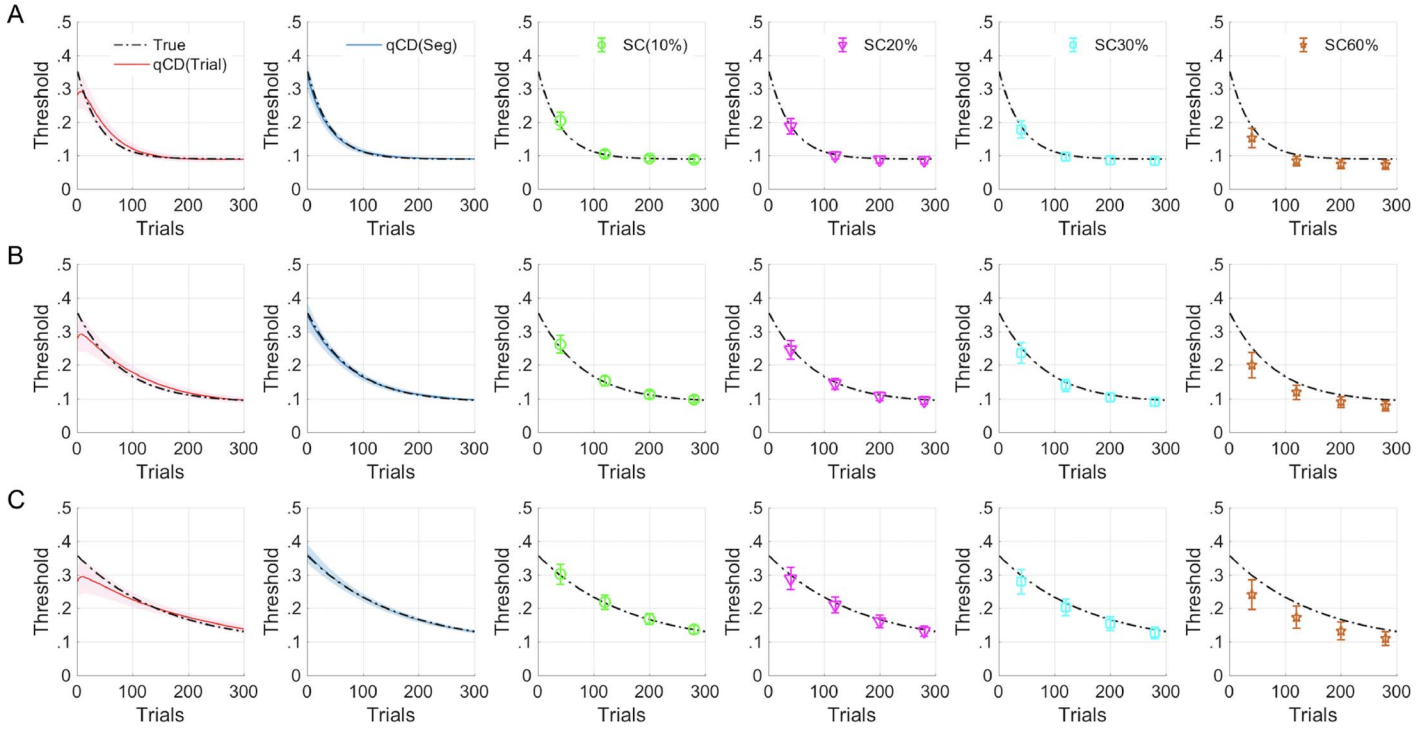

Figure 2. Estimated learning curves of the three simulated observers—(A) Observer 1, (B) Observer 2, (C) Observer 3—with a starting level at 0% from the true threshold in the first trial. Results of the first 300 trials from both the quick Change Detection and staircase methods are shown. Red and blue lines denote the estimates from trial-by-trial and post hoc segment-by-segment analyses of quick Change Detection simulation, respectively. Green circles, magenta downward-pointing triangles, cyan squares, and brown stars represent the block-by-block thresholds from the staircase method with step sizes of 10%, 20%, 30%, and 60%, respectively. The shaded area and error bars denote the standard deviation in the quick Change Detection and staircase methods, respectively. Black dash-dotted lines denote the true learning curves.

In summary, the precision of the estimated thresholds from the qCD method was much higher than that of those from the staircase methods. The *SD*s of the estimated thresholds with five starting levels are listed in Table 4; more details are provided in Supplementary Figure S7.

The 68.2% HWCIs of the estimated thresholds from the qCD method with the 0% starting level are shown in Figure 3C. Averaged across the three simulated observers, the 68.2% HWCIs of the estimated trial-by-trial threshold were 0.049, 0.042, 0.030, and 0.016 $\log_{10}$ units, respectively, after 100, 200, 400, and 800 trials; for the post hoc segment-by-segment analysis across the three simulated observers they were 0.025, 0.018, 0.013, and 0.016 $\log_{10}$ units. The 68.2% HWCIs with all five starting stimulus levels are listed in Table 5; more details are provided in Supplementary Figure S8.

Accuracy and precision of the estimated parameters

Since SC20%, SC30%, and SC60% generated estimated thresholds with much lower accuracy and precision than SC10%, we compared only SC10% with the qCD method in the following analyses. The biases and *SD*s of the estimated parameters from the qCD and staircase methods with 0% starting levels are plotted in Figure 4. For the qCD method, the bias was computed from the post hoc segment-by-segment analysis. For the staircase method, we calculated the bias and *SD* of the best-fitting parameters of the exponential model to the estimated learning curves. The biases of the estimated parameters from the qCD method were smaller than those from the staircase method, especially for simulated Observer 1, who has the fastest learning rate among the three observers.

For Observer 1, with the starting stimulus level at 0%, the biases of the estimated λ, γ, and α were 0.030, 0.028, and −0.001 $\log_{10}$ units, respectively, from the qCD method and 0.149, −0.029, and −0.004 $\log_{10}$ units from the staircase method; the *SD*s of the estimated λ, γ, and α were 0.074, 0.083, and 0.013 $\log_{10}$ units from the qCD method and 0.489, 0.213, and 0.017 $\log_{10}$ units from the staircase method. For Observers 2 and 3, with slower learning, the biases and *SD*s of the estimated parameters from the staircases were not as large as for Observer 1. However, qCD still provided more accurate (less biased) and precise estimates. In addition, different starting stimulus levels affected the accuracy of the estimated parameters from the staircase method

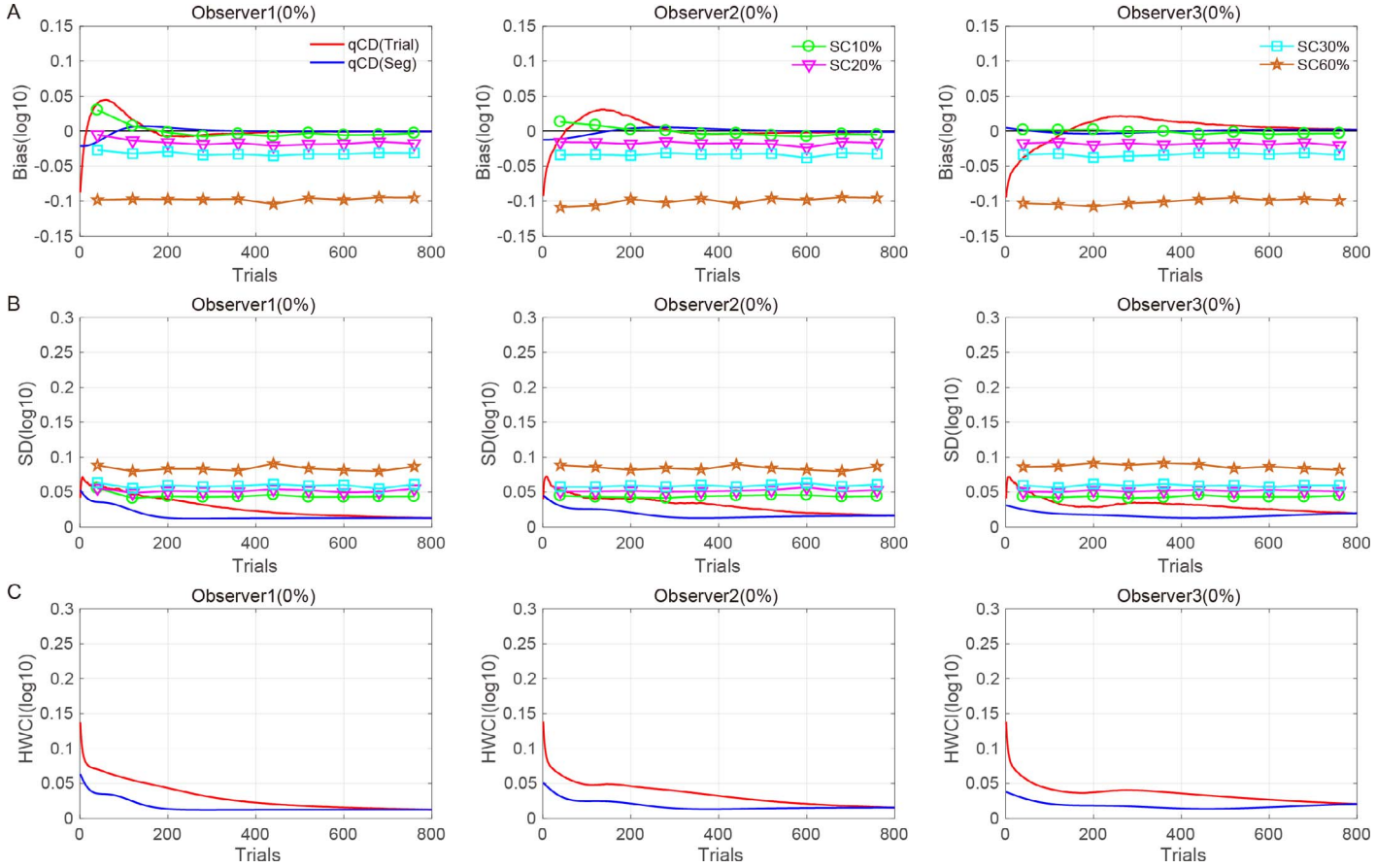

Figure 3. Comparison of the accuracy and precision of the estimated thresholds from the quick Change Detection and staircase methods. The biases (A), standard deviations (B), and half widths of 68.2% credible intervals (C) of the three simulated observers with a starting level of 0% from the true threshold in the first trial are shown in separate rows as functions of trial numbers. Red and blue lines denote the results from the trial-by-trial and post hoc analyses of quick Change Detection simulation, respectively. Green circles, magenta downward-pointing triangles, cyan squares, and brown stars represent the results from the staircase method with step sizes of 10%, 20%, 30%, and 60%, respectively.

but not the qCD method (see Supplementary Figure S9 for details).

The 68.2% HWCIs of the estimated parameters from the qCD method are shown in Figure 5; they decreased with trial number. Averaged across the three observers, the 68.2% HWCIs of the estimated λ were 0.111, 0.087, 0.072, and 0.069 $\log_{10}$ units, respectively, after 100, 200, 400, and 800 trials; for the estimated γ they were 0.196, 0.158, 0.105, and 0.074 $\log_{10}$ units; and for the estimated α they were 0.169, 0.117, 0.051, and 0.018 $\log_{10}$ units (see Supplementary Figure S10 for details on $\pm$25% and $\pm$50% starting levels). The results indicate that the qCD method can estimate the parameters of the learning curve with high precision.

	Observer 1					Observer 2					Observer 3				
Method	+50%	+25%	0%	−25%	−50%	+50%	+25%	0%	−25%	−50%	+50%	+25%	0%	−25%	−50%
qCDtrial	0.016	0.015	0.015	0.015	0.016	0.015	0.015	0.015	0.015	0.015	0.018	0.018	0.018	0.018	0.018
qCDseg	0.006	0.006	0.005	0.005	0.006	0.004	0.004	0.004	0.004	0.003	0.003	0.003	0.002	0.003	0.003
SC10%	0.006	0.010	0.011	0.090	0.004	0.006	0.007	0.007	0.004	0.006	0.005	0.004	0.003	0.029	0.007
SC20%	0.014	0.014	0.014	0.014	0.015	0.014	0.015	0.015	0.015	0.017	0.016	0.015	0.015	0.016	0.017
SC30%	0.027	0.028	0.028	0.028	0.029	0.029	0.029	0.030	0.030	0.031	0.030	0.029	0.030	0.031	0.032
SC60%	0.088	0.088	0.090	0.090	0.091	0.090	0.090	0.092	0.092	0.092	0.093	0.091	0.093	0.094	0.095

Table 3. Root mean square error of the estimated thresholds from the quick Change Detection (qCD) and staircase methods. *Notes*: qCDtrial = trial-by-trial qCD; qCDseg = post hoc segment-by-segment qCD.

	Observer 1					Observer 2					Observer 3				
Method	+50%	+25%	0%	−25%	−50%	+50%	+25%	0%	−25%	−50%	+50%	+25%	0%	−25%	−50%
qCDtrial	0.029	0.029	0.029	0.029	0.029	0.031	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032
qCDseg	0.017	0.017	0.017	0.016	0.016	0.018	0.019	0.018	0.018	0.018	0.017	0.017	0.017	0.018	0.017
SC10%	0.046	0.046	0.045	0.045	0.045	0.045	0.044	0.044	0.044	0.044	0.044	0.044	0.043	0.043	0.044
SC20%	0.052	0.053	0.052	0.052	0.053	0.052	0.052	0.052	0.052	0.052	0.051	0.051	0.052	0.052	0.052
SC30%	0.060	0.060	0.059	0.059	0.059	0.059	0.058	0.059	0.059	0.060	0.059	0.058	0.059	0.059	0.060
SC60%	0.085	0.085	0.084	0.084	0.085	0.086	0.085	0.085	0.084	0.086	0.087	0.089	0.087	0.085	0.087

Table 4. Standard deviations of the estimated thresholds in the quick Change Detection (qCD) and staircase methods. *Notes*: qCDtrial = trial-by-trial qCD; qCDseg = post hoc segment-by-segment qCD.

Estimated initial threshold and percent of improvements

Accurate and precise estimates of the initial threshold ($IT = \lambda + \alpha$) and percentage of improvements ($PI = [\lambda + \alpha]/\alpha$) are critical for understanding characteristics of perceptual learning, such as specificity, transfer, and retention. As shown in Figure 6A, the histogram of the estimated ITs from the qCD method are distributed symmetrically and tightly around the true IT (= 0.358). However, the histograms of the estimated IT from the staircase method have long tails in one direction, indicating systematic and sometimes large biases, with corresponding effects on the *SD*. For example, when the time constant was 40 trials (Observer 1), the estimated IT was 0.344 ± 0.041 ($M \pm SD$) from the qCD method and 0.847 ± 1.608 from SC10%.

The distributions of the estimated PI (Figure 6B) from the qCD method are also much narrower and closer to the true PI (= 400%) compared to those from SC10%. For example, when the time constant was 40 trials (Observer 1), the estimated PI was 384% ± 45% from the qCD method and 944% ± 1769% from SC10%. Obviously, the staircase method produced a less accurate mean and much larger *SD* for the estimated PI. These results demonstrate that the accuracy and precision of the estimated IT and PI from the qCD method were much higher than those from the staircase method. The means and *SD*s of the estimated IT and PI of the three observers with five starting levels are summarized in Table 6.

Simulation Study 2

Method

Although Simulation Study 1 found that the estimated initial threshold, which is central to the formula for the transfer index (Equation 19), was more accurate and precise from the qCD method than from the staircase method, a direct evaluation of the methods in measuring transfer/specificity is still necessary. Therefore, to further evaluate the performance of the qCD and staircase methods in assessing the transfer

Observer	Starting level	Trial 1	Trial 100	Trial 200	Trial 400	Trial 800	Average
Observer 1	+50%	0.150; 0.065	0.058; 0.029	0.044; 0.014	0.023; 0.012	0.013; 0.013	0.031; 0.017
	+25%	0.147; 0.065	0.058; 0.029	0.043; 0.014	0.023; 0.013	0.013; 0.013	0.031; 0.017
	0%	0.137; 0.063	0.058; 0.029	0.043; 0.014	0.023; 0.013	0.013; 0.013	0.031; 0.017
	−25%	0.133; 0.064	0.059; 0.029	0.043; 0.014	0.023; 0.013	0.013; 0.013	0.031; 0.017
	−50%	0.141; 0.064	0.058; 0.029	0.044; 0.014	0.023; 0.012	0.013; 0.013	0.031; 0.017
Observer 2	+50%	0.150; 0.051	0.048; 0.025	0.046; 0.021	0.032; 0.013	0.016; 0.016	0.035; 0.018
	+25%	0.147; 0.051	0.048; 0.025	0.046; 0.021	0.032; 0.013	0.016; 0.016	0.035; 0.018
	0%	0.138; 0.051	0.048; 0.025	0.046; 0.021	0.032; 0.013	0.016; 0.016	0.035; 0.018
	−25%	0.133; 0.050	0.048; 0.025	0.046; 0.021	0.032; 0.013	0.015; 0.015	0.035; 0.018
	−50%	0.140; 0.051	0.048; 0.025	0.046; 0.021	0.032; 0.013	0.015; 0.015	0.035; 0.018
Observer 3	+50%	0.150; 0.038	0.043; 0.021	0.037; 0.019	0.036; 0.014	0.021; 0.021	0.036; 0.018
	+25%	0.148; 0.038	0.042; 0.021	0.037; 0.019	0.036; 0.014	0.021; 0.021	0.036; 0.018
	0%	0.138; 0.038	0.042; 0.021	0.037; 0.018	0.036; 0.014	0.021; 0.021	0.036; 0.018
	−25%	0.148; 0.038	0.042; 0.021	0.037; 0.019	0.036; 0.014	0.021; 0.021	0.036; 0.018
	−50%	0.140; 0.038	0.042; 0.021	0.037; 0.019	0.036; 0.014	0.021; 0.021	0.036; 0.018

Table 5. Half widths of 68.2% credible intervals of the estimated thresholds (in $\log_{10}$ units) from the quick Change Detection (qCD) method: trial by trial and post hoc segment by segment.

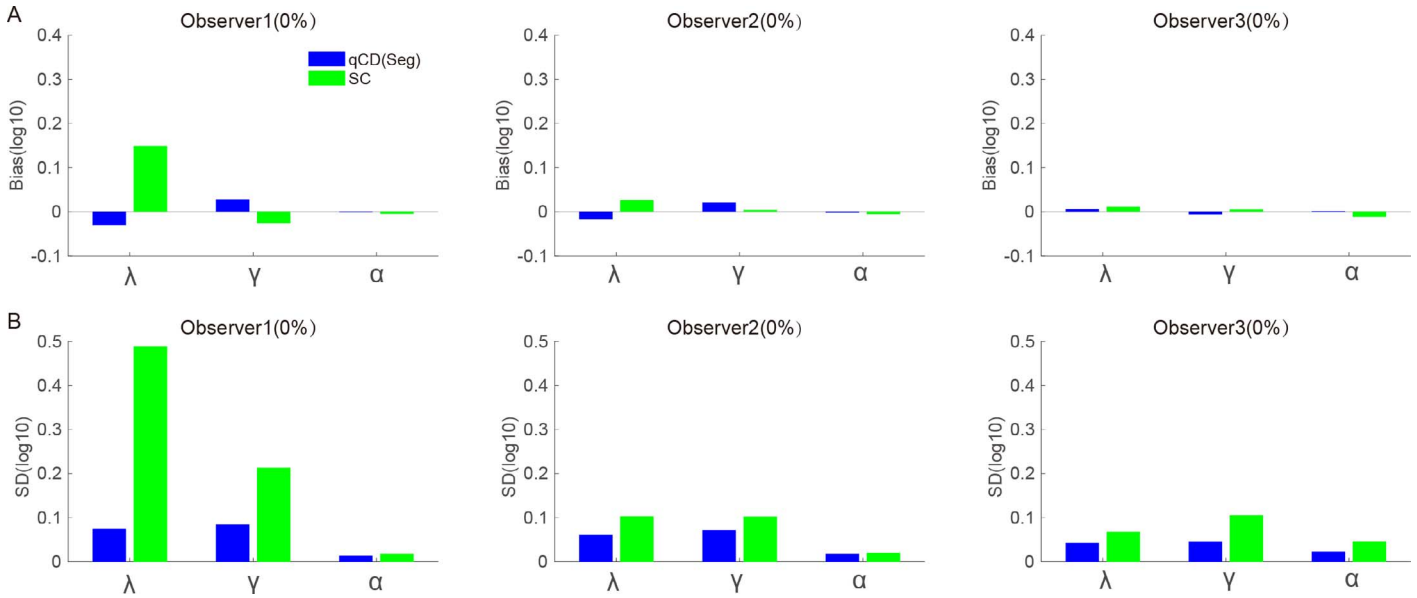

Figure 4. The biases (A) and standard deviations (B) of the estimated parameters (λ, γ, and α) of the three simulated observers from the post hoc segment-by-segment quick Change Detection (blue) and SC10% staircase (green) methods.

index in a 2AFC task, we extended the simulations in Simulation Study 1 by adding a transfer phase to each of the simulated observers. Specifically, each simulated observer kept the same time constant and asymptotic performance level but had only half threshold reduction in the transfer phase (Table 7).

We conducted 1,000 simulated transfer runs for each observer with each method. Each run consisted of 800 trials of the qCD method and 800 trials of the 3-down/1-up staircase method (Cornsweet, 1962) with a 10% step size and 80-trial block size. Both methods were simulated with five starting stimulus levels (+50%, +25%, 0%, −25%, and −50% from the true initial threshold in the training phase). We used the same starting stimulus levels in the training and transfer phases (Donovan & Carrasco, 2015; Jeter, Dosher, Petrov, & Lu, 2009; Liang et al., 2015; Xiao et al., 2008; Zhang, Hou, et al., 2018). The qCD and staircase procedures were run separately and independently, and thus never shared any information. In other words, the terminology "training" and "transfer" in the current study is purely nominal. The detailed settings of Simulation Study 2 were the same as those of Simulation Study 1.

The transfer index (TI) is defined as

$$\mathrm{TI} = 1 - \frac{T_{B1} - T_{Bend}}{T_{A1} - T_{Aend}}, \quad (19)$$

where T_{A1} and T_{Aend} are the estimated thresholds in the first and last trials of the training, and T_{B1} and T_{Bend} are the estimated thresholds for the first and last trials of the transfer. For the qCD method, TI was computed using the estimated thresholds from the post hoc segment-by-segment analysis. For the staircase method, the TIs were obtained from the best-fitting exponential function to the estimated block-by-block thresholds in each run. As shown in Table 7, the true TI of all three simulated observers was 50%.

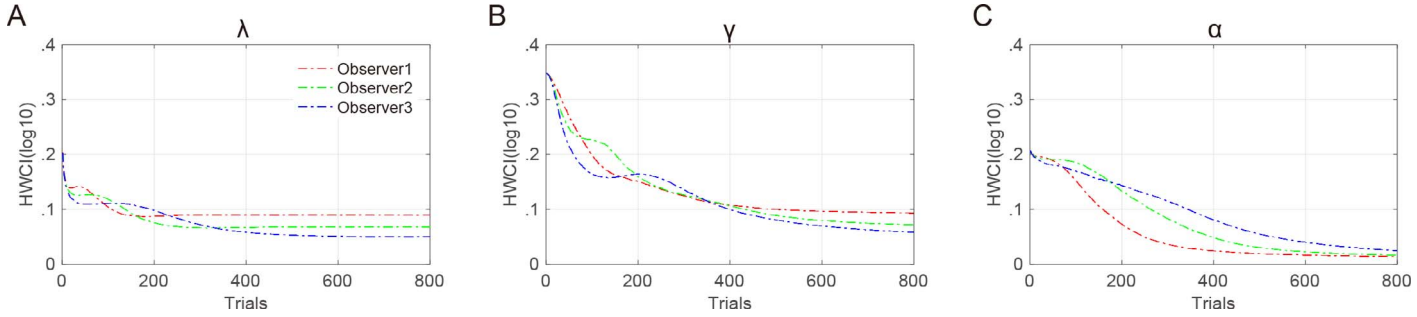

Figure 5. Half widths of 68.2% credible intervals of the estimated λ (A), γ (B), and α (C) of the three simulated observers with starting levels at 0% from the true threshold in the first trial in the trial-by-trial quick Change Detection simulation. Red, green, and blue dash-dotted lines denote Observers 1, 2, and 3, respectively.

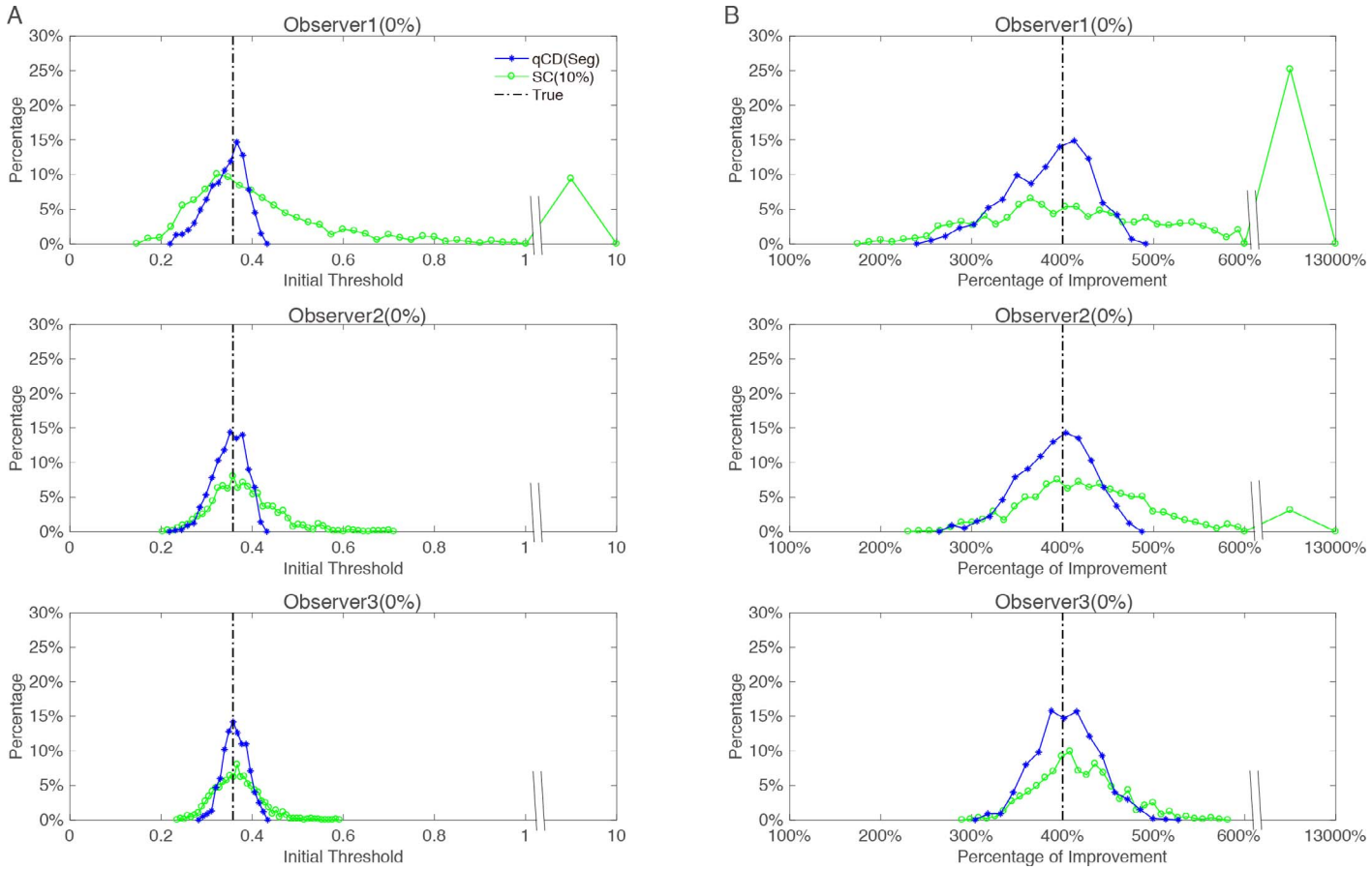

Figure 6. Distributions of the initial threshold (A) and percentage of improvements (B) from the quick Change Detection and SC10% staircase methods with the 0% starting level. Blue asterisks and green circles denote results from the post hoc segment-by-segment quick Change Detection and staircase methods, respectively. Black dash-dotted lines denote the true values. Results from the three simulated observers are shown in separate rows.

Results

Staircase convergence

Again, the estimated threshold in each measurement block of the staircase was calculated by averaging stimulus levels at an even number of reversals after deleting the first four or five of them. Table 8 shows that the percentage of staircases in the first block of the transfer that had more than five reversals was 100% in all but two conditions for Observer 1.

		Observer 1		Observer 2		Observer 3	
Measure	Starting level	qCDseg	SC10%	qCDseg	SC10%	qCDseg	SC10%
IT	+50%	0.341 (0.041)	0.799 (1.609)	0.351 (0.036)	0.380 (0.079)	0.364 (0.026)	0.375 (0.053)
	+25%	0.340 (0.041)	0.901 (1.737)	0.351 (0.037)	0.387 (0.076)	0.365 (0.027)	0.372 (0.051)
	0%	0.344 (0.041)	0.847 (1.608)	0.350 (0.035)	0.382 (0.071)	0.363 (0.026)	0.366 (0.048)
	−25%	0.344 (0.040)	0.798 (1.534)	0.352 (0.037)	0.369 (0.123)	0.363 (0.026)	0.355 (0.046)
	−50%	0.340 (0.042)	0.766 (1.516)	0.352 (0.042)	0.344 (0.069)	0.364 (0.026)	0.342 (0.044)
PI	+50%	381% (46%)	894% (1,787%)	394% (40%)	429% (84%)	405% (35%)	426% (51%)
	+25%	381% (46%)	1010% (1,949%)	393% (41%)	435% (81%)	406% (33%)	431% (206%)
	0%	384% (45%)	944% (1,769%)	392% (39%)	431% (76%)	405% (34%)	418% (46%)
	−25%	384% (45%)	894% (1,711%)	395% (40%)	417% (131%)	405% (34%)	411% (48%)
	−50%	381% (47%)	852% (1,662%)	396% (40%)	390% (71%)	406% (33%)	405% (54%)

Table 6. Means and standard deviations of initial threshold (IT) and percentage of improvements (PI) in the quick Change Detection (qCD) and staircase methods. *Notes*: qCDseg = post hoc segment-by-segment qCD.

	Observer 1		Observer 2		Observer 3	
Parameter	Training	Transfer	Training	Transfer	Training	Transfer
λ	0.2685	0.1343	0.2685	0.1343	0.2685	0.1343
γ	40	40	80	80	160	160
α	0.0895	0.0895	0.0895	0.0895	0.0895	0.0895
TI		50%		50%		50%

Table 7. Parameters of the three simulated observers in the training and transfer phases of Simulation Study 2. *Notes*: TI = transfer index.

Estimated learning curves in the transfer phase

For brevity, in the main text we present only the results from the qCD method and the staircase methods in the transfer phases using the 0% starting level. Visual inspection suggests that the estimated post hoc segment-by-segment thresholds from the qCD method were very close to the true thresholds, and closer than the block-by-block thresholds from the staircase method (Figure 7). In addition, the estimated thresholds from the qCD method were more precise (less variable) than those from the staircase method. We quantify these observations next. The estimated thresholds with ±25% and ±50% starting levels are shown in Supplementary Figures S11 through S13.

Accuracy and precision of the estimated thresholds in the transfer phase

The biases of the estimated thresholds from the qCD and staircase methods with the 0% starting level in the transfer phase are plotted in Figure 8A. The RMSEs of the estimated thresholds from the qCD trial-by-trial, qCD post hoc segment-by-segment, and SC10% were 0.013, 0.003, and 0.006 $\log_{10}$ units, respectively, for Observer 1; 0.009, 0.003, and 0.005 $\log_{10}$ units for Observer 2; and 0.015, 0.005, and 0.004 $\log_{10}$ units for Observer 3. The qCD method was much more accurate than the staircase method when the learning rate was more rapid. The RMSEs of the estimated thresholds with starting levels of ±25% and ±50% are listed in Supplementary Table S2; more details are provided in Supplementary Figure S14.

The *SD*s of the estimated thresholds from the qCD and staircase methods with the 0% starting level in the transfer phase are shown in Figure 8B. The *SD*s of the estimated thresholds from the qCD trial-by-trial, qCD post hoc segment-by-segment, and SC10% were 0.027, 0.016, and 0.045 $\log_{10}$ units, respectively, for Observer 1; 0.030, 0.017, and 0.044 $\log_{10}$ units for Observer 2; and 0.031, 0.018, and 0.044 $\log_{10}$ units for Observer 3. For the staircase methods, the *SD* increased with step size. The *SD*s of the estimated thresholds from the qCD method were always considerably smaller than those from the staircase methods. In summary, the precision of the estimated thresholds from the qCD method was much higher than from the staircase methods. More details are provided in Supplementary Table S3 and Supplementary Figure S15.

Observer	+50%	+25%	0%	−25%	−50%
Observer 1	99.8%	99.8%	100%	100%	100%
Observer 2	100%	100%	100%	100%	100%
Observer 3	100%	100%	100%	100%	100%

Table 8. Percentage of staircases with more than five reversals in the first block of learning, by starting level.

The 68.2% HWCIs of the estimated thresholds from the qCD method with the 0% starting level in the transfer phase are shown in Figure 8C. Averaged across the three simulated observers, the 68.2% HWCIs of the estimated trial-by-trial threshold were 0.048, 0.040, 0.028, and 0.015 $\log_{10}$ units, respectively, after 100, 200, 400, and 800 trials; from the post hoc segment-by-segment analysis they were 0.024, 0.017, 0.013, and 0.015 $\log_{10}$ units. More details are provided in Supplementary Figure S16.

Accuracy and precision of the estimated parameters in the transfer phase

The biases and *SD*s of the estimated parameters from the qCD and staircase methods with 0% starting level (relative to the true threshold in the beginning of the training phase) in the transfer phase are plotted in Figure 9. The biases of the estimated parameters from the qCD method were smaller than those from the staircase method, especially for simulated Observer 1, who has the fastest learning rate among the three observers. For Observer 1, with the starting stimulus level at 0%, the biases of the estimated λ, γ, and α were 0.042, −0.009, and −0.001 $\log_{10}$ units, respectively, from the qCD method and 0.302, −0.042, and −0.028 $\log_{10}$ units from the staircase method; the *SD*s were 0.087, 0.107, and 0.013 $\log_{10}$ units from the qCD method and 0.978, 0.453, and 0.470 $\log_{10}$ units from the staircase methods. For Observers 2 and 3, with slower learning, the biases and *SD*s of the estimated parameters from the staircases were not as large as for Observer 1. However, qCD still provided more accurate and precise estimates. In addition, different

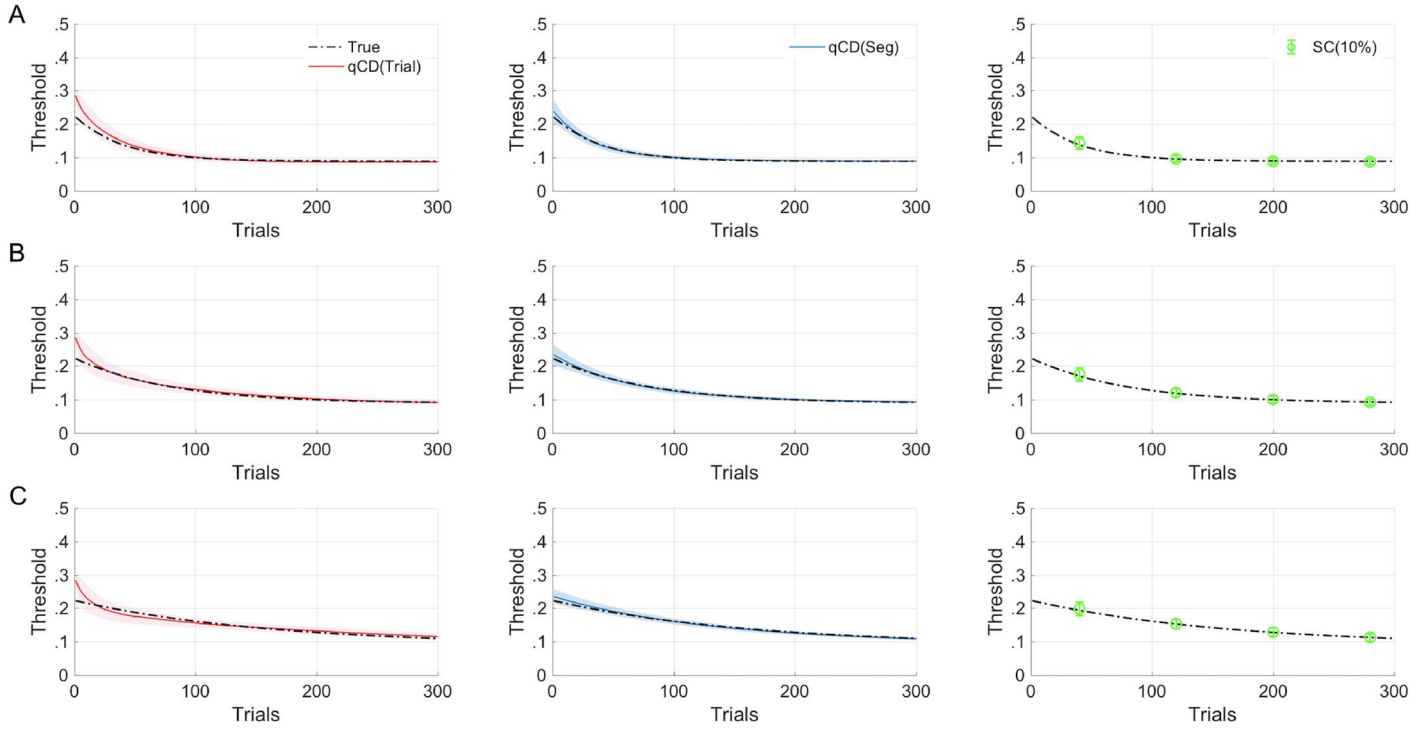

Figure 7. Estimated learning curves in the transfer phase. Results from the first 300 trials of the three simulated observers are shown in separate rows. Red and blue lines denote the estimates from trial-by-trial and post hoc segment-by-segment analyses of quick Change Detection simulation, respectively. Green circles represent the block-by-block threshold from the SC10% staircase method. The shaded areas and error bars denote the standard deviations in the quick Change Detection and staircase methods, respectively. Black dash-dotted lines represent the true learning curves.

starting stimulus levels affected the accuracy of the estimated parameters from the staircase method but not the qCD method (see Supplementary Figure S17 for details).

The 68.2% HWCIs of the estimated parameters from the qCD method in the transfer phase are shown in Figure 10; they decreased with trial number. Averaged across the three observers, the 68.2% HWCIs of the estimated λ were 0.125, 0.109, 0.099, and 0.095 $\log_{10}$ units, respectively, after 100, 200, 400, and 800 trials; for the estimated γ they were 0.276, 0.242, 0.150, and 0.109 $\log_{10}$ units; and for the estimated α they were 0.151, 0.098, 0.040, and 0.017 $\log_{10}$ units (see Supplementary Figure S18 for details on $\pm$25% and $\pm$50% starting levels). The results indicate that the qCD method can estimate the parameters of the learning curve with high precision.

Estimated IT and PI

The histogram of the estimated ITs and PIs from the qCD method is distributed symmetrically and tightly around the true values (IT = 0.224, PI = 250%), but that from the staircase method is not (see Figure 11), indicating that the qCD method produced estimated ITs and PIs with smaller biases and *SD*s. Consistent with the results of Simulation Study 1, the qCD method produced estimated ITs and PIs with higher accuracy and precision than the staircase method. The means and *SD*s of the estimated ITs and PIs of the three observers with five starting levels are summarized in Supplementary Table S4.

Estimated TI

The histogram of the estimated TIs of the three observers with the 0% starting level is shown in Figure 12. The histograms of the estimated TIs from the qCD and staircase method are both distributed symmetrically around the true value (= 50%), but the staircase method produced tighter distributions for Observers 1 and 2 (see Supplementary Figure S19 for starting levels of $\pm$25% and $\pm$50%). The estimated TIs from the qCD and staircase methods, averaged over starting levels, were respectively 39% $\pm$ 17% ($M \pm SD$) and $-$121% $\pm$ 491% for Observer 1, 43% $\pm$ 14% and 25% $\pm$ 219% for Observer 2, and 47% $\pm$ 9% and 46% $\pm$ 48% for Observer 3. The results indicate that the qCD method provided more accurate and precise measures of transfer than the staircase method, especially when the learning is more rapid.

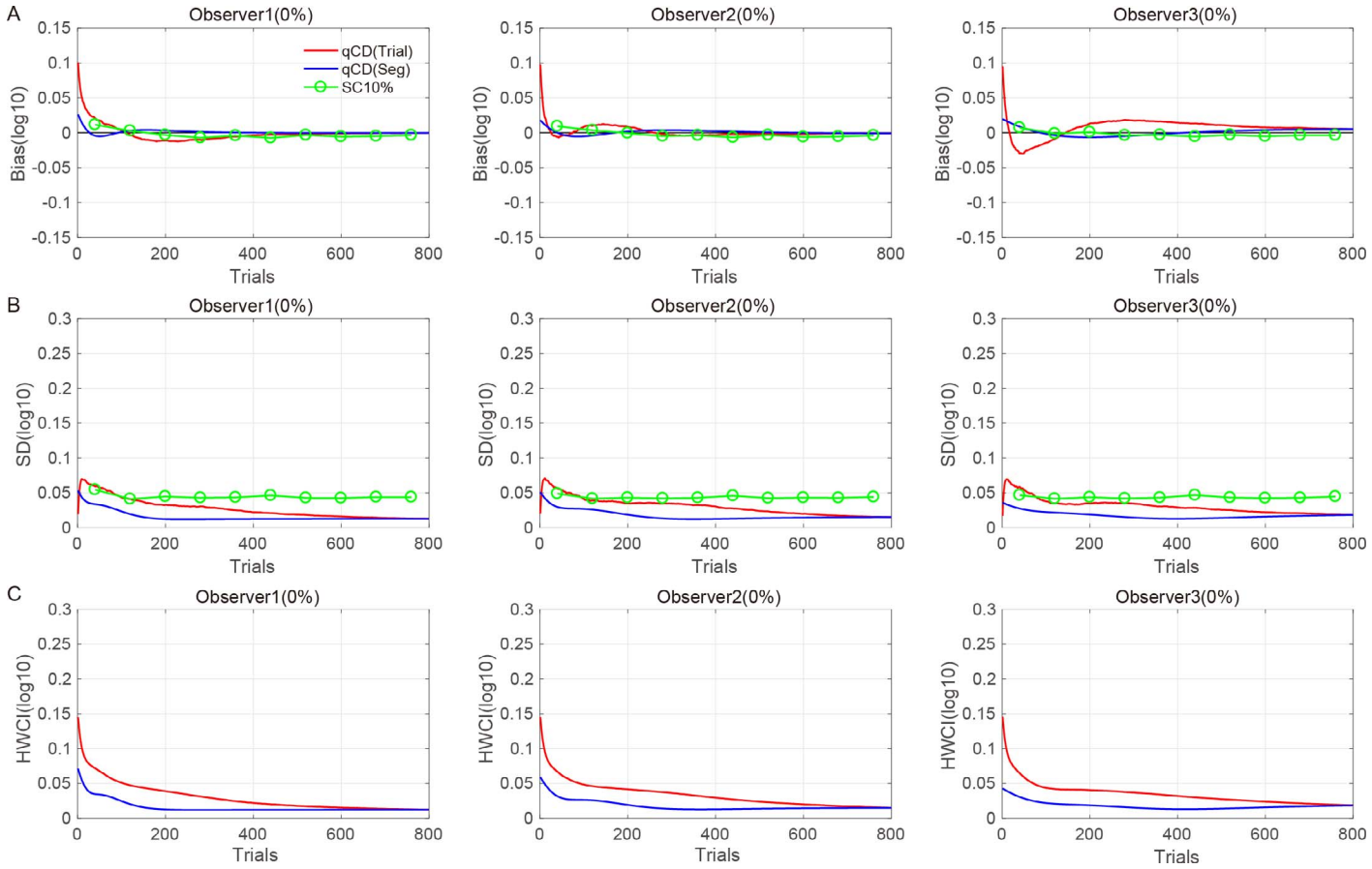

Figure 8. Comparison of the accuracy and precision of the estimated thresholds in the transfer phase. The biases (A), standard deviations (B), and half widths of 68.2% credible intervals (C) of the three simulated observers with starting level at 0% from the true threshold in the first trial are shown in separate rows as functions of trial numbers. Red and blue lines denote the results from the trial-by-trial and post hoc analyses of quick Change Detection simulations, respectively. Green circles represent the results from the SC10% staircase method.

Discussion

In this study, we systematically examined the performance of the 3-down/1-up staircase and qCD methods in measuring the detailed time course of perceptual learning and transfer in a 2AFC task. The staircase method generated biased and imprecise estimations of the learning curve and transfer index, while the qCD method provided accurate and precise estimations. The staircase method did not converge when the block size was 40 trials. Larger block sizes (e.g., 160 trials) resulted in fewer block thresholds and therefore inaccurate and imprecise estimates of the parameters of the learning curves. Staircases with 1% and 5% step sizes failed to generate more than five reversals half of the time. Staircases with 10%, 20%, 30%, and 60% step sizes produced imprecise and biased estimation of thresholds and parameters of the learning curves. The staircase method could not provide an accurate and precise estimate of the transfer index even when the starting stimulus level was set optimally. The faster the learning was, the worse the staircase method performed. And the qCD method characterized the time course of perceptual learning and transfer more accurately, precisely, and efficiently than the staircase method.

Zhang, Zhao, et al. (2018, 2019) implemented and tested the qCD method in assessing the learning curve in a 4AFC global motion-direction identification task both in simulations and in a psychophysical experiment. In their psychophysical experiment, the learning curves estimated from the qCD and staircase methods matched quite well for this relatively slow learning task. However, the qCD method still provided a more precise assessment of the learning curve. In their simulation experiment, with any one of three starting stimulus levels (+25%, 0%, and −25% from the true threshold), the estimated learning curves from both the trial-by-trial and post hoc segment-by-segment analyses of the qCD method were always more accurate and precise than those obtained from the staircase method using either 80- or 160-trial blocks. Going beyond

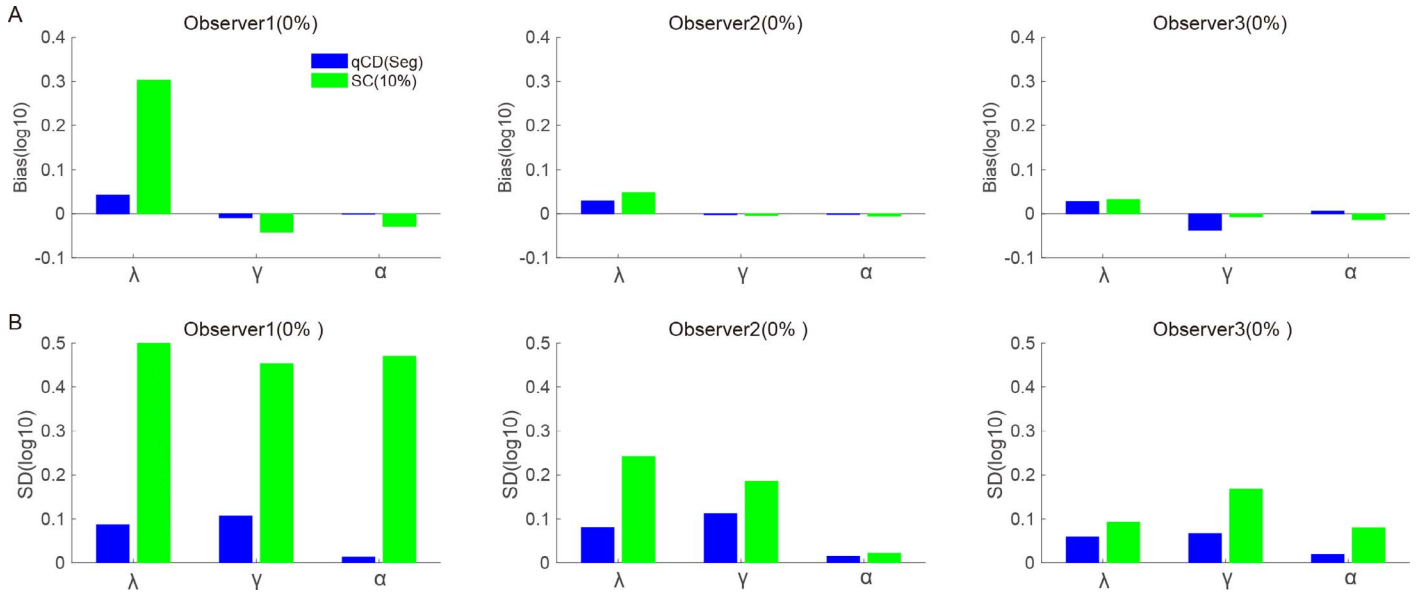

Figure 9. The biases (A) and standard deviations (B) of the estimated parameters (λ, γ, and α) in the transfer phase. Results from the post hoc segment-by-segment quick Change Detection analyses are shown in blue. Results from the SC10% staircase are shown in green.

previous work, the current simulation studies systematically evaluated the performance of the 3-down/1-up staircase method with six step sizes, five starting levels, and three block sizes on accessing the detailed time course and transfer of perceptual learning in a 2AFC task, which is the most widely used procedure in perceptual-learning studies (Jogan & Stocker, 2014; Kingdom & Prins, 2010; Vancleef et al., 2018).

Step sizes of the staircase method were manipulated at six levels (1%, 5%, 10%, 20%, 30%, and 60%) to examine which step size could best track the detailed time course of perceptual learning. Smaller step sizes (1% and 5%) failed to generate reliable threshold estimates, because the staircase method does not generate enough reversals. Larger step sizes (20%, 30%, and 60%) produced block threshold estimates with lower accuracy and precision compared to the 10% step size. Although the commonly used 10% step size is the best option for the staircase method, it still produced estimates with substantial bias and *SD* for the estimated learning curve parameters in some cases. Furthermore, when learning is more rapid, the staircase method is less able to track the detailed time course of perceptual learning. This is not surprising, since there is an insufficient number of block measurements to track reductions in threshold, especially in the early rapid-changing phase of learning.

In the current study, the starting levels (initial threshold levels tested) also influenced the performance of the staircase method in estimating learning curves. The biases and *SD*s of the estimated parameters from the staircase method depended more on the starting levels, while the qCD method was much less dependent on the deviation of the starting value from the true value. The staircase method performed worse when the starting levels were set nonoptimally.

The staircase method generated biased and imprecise threshold estimates because a constant threshold is

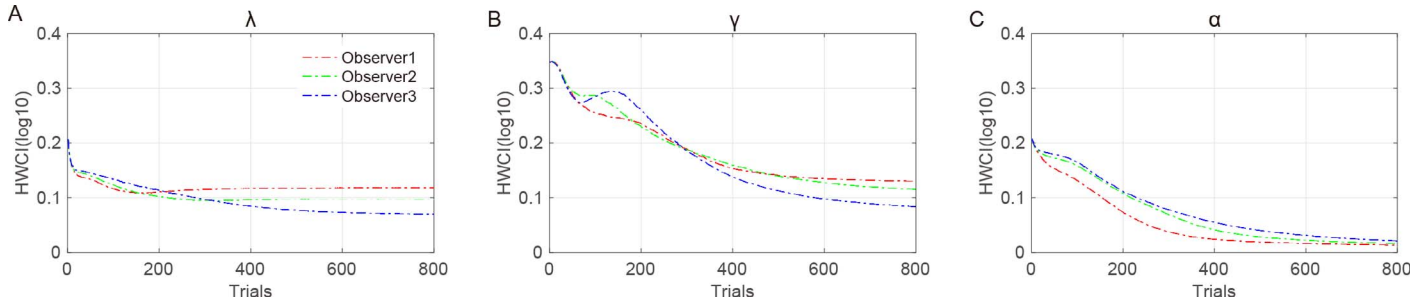

Figure 10. Half widths of 68.2% credible intervals of the estimated λ (A), γ (B), and α (C) from the quick Change Detection method in the transfer phase. Red, green, and blue dash-dotted lines denote Observers 1, 2, and 3, respectively.

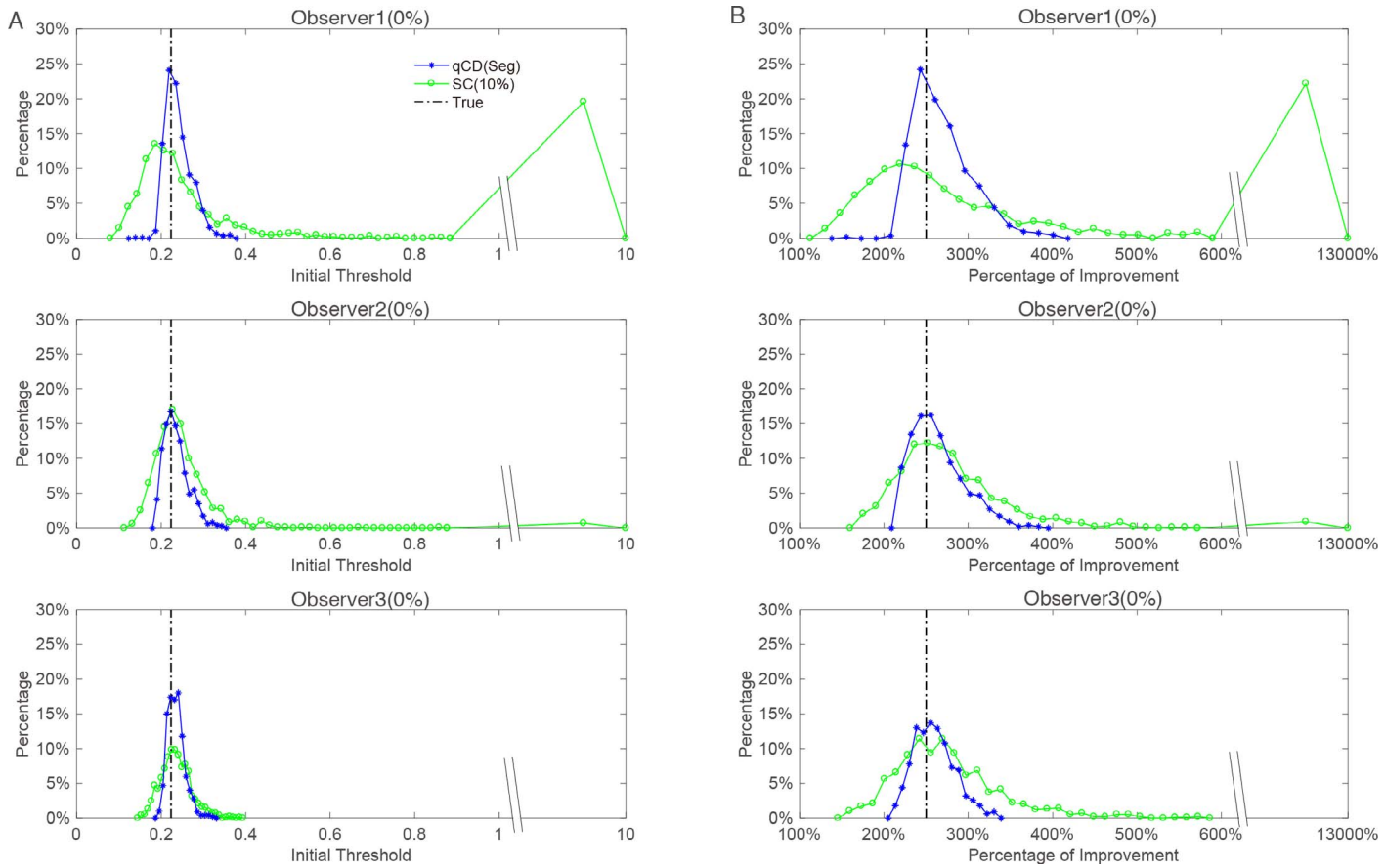

Figure 11. Distributions of the estimated initial threshold (A) and percentage of improvements (B) in the transfer phase from the quick Change Detection and SC10% staircase methods. Blue asterisks and green circles denote results from the post hoc segment-by-segment quick Change Detection and staircase methods, respectively. Black dash-dotted lines denote the true values. Results from the three observers are shown in separate rows.

assumed within each block of trials, whereas the perceptual sensitivity during perceptual learning changes continuously (Lu et al., 2011; Mazur & Hastie, 1978; Petrov et al., 2005), even within each measurement block and especially in the early phase of learning (Badiru, 1992; Dosher & Lu, 2007; Heathcote et al., 2000). On the other hand, the qCD method parameterizes the learning curve as an exponential function

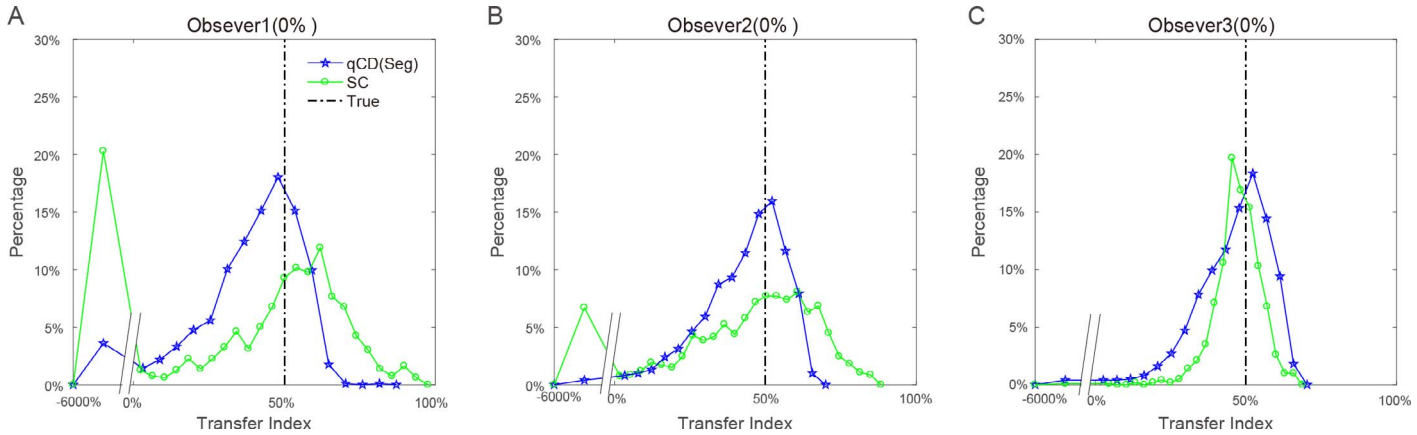

Figure 12. Distributions of the transfer index of three simulated observers—(A) Observer 1, (B) Observer 2, (C) Observer 3—with a starting level at 0% from the true threshold in the first trial. Blue and green symbols denote the results from post hoc segment-by-segment quick Change Detection and SC10% staircase methods. Black dash-dotted lines denote the true values.

with joint probability distributions of all three parameters. Based on the adaptive Bayesian framework (Kontsevich & Tyler, 1999; Lesmes, Jeon, Lu, & Dosher, 2006; Lesmes et al., 2010; Watson & Pelli, 1983), the qCD method selects the next test stimulus to optimize the expected information gain trial by trial. By continuously updating the posterior distribution trial by trial, the qCD method provides a detailed trial-by-trial estimate of perceptual sensitivity that tracks the detailed time course of perceptual-sensitivity change. Our simulations show that the estimated learning curves and their parameters from the qCD method were very accurate and precise with the 2AFC task tested here regardless of the deviations in starting levels.

Although specificity is considered to be the hallmark of perceptual learning, a growing number of studies have found that learning can transfer and that the degree of transfer depends critically on the difficulty (Ahissar & Hochstein, 1997; Z. Liu, 1995; Z. Liu, 1999) or precision (Jeter et al., 2009) of the training or transfer tasks, training duration (Jeter, Dosher, Liu, & Lu, 2010), and training procedure (J. Huang, Liang, Zhou, & Liu, 2017; Liang et al., 2015; Tartaglia, Bamert, Mast, & Herzog, 2009; Xiao et al., 2008). Accurate and precise estimates of the initial and final thresholds in the learning phase and the initial threshold in the transfer phase are critical for computing the transfer index (Ahissar & Hochstein, 1997; Dosher & Lu, 2007; Jeter et al., 2010; J. Liu, Lu, & Dosher, 2012). In Simulation Study 1, the estimated initial threshold from the staircase method was quite biased and imprecise, while the qCD method yielded estimated initial thresholds with high accuracy and precision. In Simulation Study 2, we found that the estimated transfer index from the staircase method was much less accurate and precise than from the qCD method, especially when the learning was rapid. The qCD method performed well with all starting stimulus levels and learning rates. In these stimulations, we assumed that the learning rates were the same in the learning and transfer phases. Theoretically, transfer of perceptual learning could lead to a lower initial threshold or faster learning (Z. Liu & Weinshall, 2000). This could be even more challenging for the staircase method to estimate accurately.

Although the priors in the current study were uniform rather than the weakly informative secant functions used by Zhang, Zhao, et al. (2018, 2019), the qCD procedure still produced estimates with relatively high accuracy and precision. Our qCD procedure is also superior to that of Kattner, Cochrane, and Green (2017). Although their procedure can provide an estimated continuous trial-by-trial learning curve that is more detailed than the block-by-block analysis typically used in perceptual-learning studies with the staircase method, the selection of test stimuli during learning was not optimized during data collection. As a result, the new fitting procedure significantly improved the confidence intervals of the estimated thresholds in the late phase of the learning curve, but did not substantially benefit estimates in the early phases of learning.

Per the request of an anonymous reviewer, we rescored the simulated data from the staircase method using the Bayesian estimation component of the qCD method. The data of all three simulated observers (training part only) from the staircase method with three starting levels (0%, +50%, and −50% above the true initial threshold) and two step sizes (10% and 5%) were rescored using the trial-by-trial and post hoc segment-by-segment qCD procedures. For the data obtained with both step sizes, the average accuracy and precision (RMSE, HWCI, and *SD*) of the rescored thresholds were comparable to those of the direct estimates with the qCD method, except that the accuracy of the rescored thresholds of the simulated observer with the fastest learning rate, 5% step size, and +50% starting level were worse than those obtained with the qCD method directly (RMSE: 0.024 vs. 0.006 $\log_{10}$ units; Supplementary Table S5; Supplementary Figures S20 through S23). These results are different from those of Zhao et al. (2019), who found that even after rescoring with the qCD method, the accuracy and precision of the estimated thresholds from the staircase methods were worse than those obtained directly with the qCD method. The difference between the two studies is that Zhao et al. investigated faster perceptual-sensitivity changes than we did (time constants of 20 vs. 40 trials).

To further evaluate the effects of rescoring, we performed additional simulations on the training part of simulated Observer 1 (time constant = 40 trials) in 4AFC and 8AFC tasks and measured the performance with the staircase (step size = 10%) and qCD methods. In the 4AFC task, the starting levels were +22%, +84%, and −36% relative to the true initial threshold. In the 8AFC task, the starting levels were 0%, +50%, and −50% relative to the true initial threshold. The data obtained from the staircase method were rescored using the qCD method. We found that the average accuracy and precision of the rescored thresholds were comparable to those of the direct estimates from the qCD method, except when the starting level was much higher than the true initial threshold. For example, in the post hoc segment-by-segment analysis, the RMSEs of the estimated learning curve from the qCD and rescored staircase data were 0.002 and 0.005 $\log_{10}$ units, respectively, in the 4AFC task and 0.001 and 0.008 $\log_{10}$ units in the 8AFC task (Supplementary Table S6).

The qCD method exhibited some advantages in estimating thresholds in the early phase of learning. Following the segment-by-segment analysis, in the 4AFC task the biases of the estimated threshold from the qCD method were −0.008, −0.009, and −0.009, respectively, after 1, 10, and 20 trials, and for the rescored staircase data they were −0.027, −0.024, and −0.021 (all in $\log_{10}$ units); in the 8AFC task they were −0.003, −0.004, and −0.004 for the qCD method and −0.043, −0.038, and −0.032 for the rescored staircase data (Supplementary Figure S24). The same pattern of results was also true for the HWCIs of estimated thresholds. In the 4AFC task, the HWCIs of the estimated threshold from the qCD method were 0.050, 0.040, and 0.032 $\log_{10}$ units, respectively, after 1, 10, and 20 trials, and for the rescored staircase data they were 0.063, 0.052, and 0.042 (Supplementary Figure S25). To examine the effect of the number of alternatives on the efficiency of the qCD method, we also evaluated the efficiency of the qCD method in 2AFC, 4AFC, and 8AFC tasks for Observer 1 (time constant = 40 trials). Defining efficiency as the number of trials for the procedure to reach criterion accuracy (RMSE) and precision (*SD*) levels (Zhao et al., 2019), we found that the qCD procedure required 93, 84, and 76 trials to reach 0.02 $\log_{10}$-unit accuracy in 2AFC, 4AFC, and 8AFC tasks, respectively, and 677, 379, and 291 trials to reach 0.02 $\log_{10}$-unit precision (see Supplementary Figures S26 and S27). Putting Zhao et al. (2019) and the current study together, we conclude the following: Rescoring the staircase-method data with the qCD method could lead to comparable accuracy and precision with direct qCD estimates when perceptual sensitivity changes relatively slowly, but not when it changes fast; the qCD method holds advantages in early phase of learning; and the qCD method is more efficient with more alternatives in forced-choice tasks.

We focused on the staircase method in this study because it is the most widely used procedure in assessing effects of perceptual learning. Here we briefly discuss the difference between qCD and other adaptive procedures. QUEST is a classic adaptive Bayesian testing procedure for estimating threshold from a sequence of psychophysical trials (Watson & Pelli, 1979, 1983). It assumes a single stimulus dimension and two possible trial outcomes, and estimates a single psychometric-function parameter that is defined on the stimulus dimension. Recently, QUEST+ extended QUEST to allow estimation of multiple parameters of the psychometric function, more stimulus dimensions, and more trial outcomes, as well as any form of the psychometric function and flexible sampling of stimulus and parameter dimensions (Watson, 2017). However, the qCD method is quite different from QUEST. Like many existing adaptive psychophysical procedures, QUEST assumes that the threshold does not change over time or from trial to trial. The procedures are not designed to measure changing thresholds or perceptual sensitivity in situations such as dark adaptation and perceptual learning. Although it can be extended to assess the time-course perceptual-sensitivity change, the current implementation and applications of QUEST+ have focused on thresholds that do not change over time (Watson, 2017). Additional work is necessary to extend and evaluate QUEST+ in measuring the time course of perceptual-sensitivity change.

Zhao et al. (2019) compared the performance of the qCD and quick Forced-Choice (qFC; Lesmes et al., 2015) methods in measuring the time course of dark adaptation. The qFC method, similar to QUEST and QUEST+, belongs to a family of novel Bayesian adaptive methods that are designed to estimate thresholds in yes/no and forced-choice tasks. Like QUEST and current implementations of QUEST+, the qFC method does not explicitly model perceptual-sensitivity change over time. Simulations showed that the accuracy and precision of the estimated dark-adaptation curve after one qCD run (RMSE = 0.002; HWCI = 0.016; *SD* = 0.020; all in $\log_{10}$ units) were much higher than those obtained by 10 runs of the qFC procedure (RMSE = 0.020; HWCI = 0.032; *SD* = 0.031).

In conclusion, we have systematically examined the performance of the 3-down/1-up staircase and qCD methods in measuring the detailed time course of perceptual learning and transfer in a 2AFC task. Our simulations demonstrated that the qCD method provides far more accurate and precise estimations of the learning curve and the transfer index than the staircase methods.

Keywords: quick Change Detection, perceptual learning, staircase, transfer, adaptive testing

Acknowledgments

This research was supported by the National Eye Institute Grants EY017491 and EY021553.

Commercial relationships: Z-LL has an equity interest in Adaptive Sensory Technology, Inc., San Diego, CA, USA (I); YZ and Z-LL have an intellectual-property interest in methods for measuring behavioral changes of processes, PCT/US18/21944 (P).
Corresponding author: Zhong-Lin Lu.
Email: lu.535@osu.edu.
Address: Laboratory of Brain Processes (LOBES), Department of Psychology, The Ohio State University, Columbus, OH, USA.

References

Aberg, K. C., & Herzog, M. H. (2012). Different types of feedback change decision criterion and sensitivity differently in perceptual learning. *Journal of Vision*, *12*(3), 3–3.

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*(6631), 401–406.

Andersen, G. J., Ni, R., Bower, J. D., & Watanabe, T. (2010). Perceptual learning, aging, and improved visual performance in early stages of visual processing. *Journal of Vision*, *10*(13), 4.

Badiru, A. B. (1992). Computational survey of univariate and multivariate learning curve models. *IEEE Transactions on Engineering Management*, *39*(2), 176–188.

Ball, K., & Sekuler, R. (1982). A specific and enduring improvement in visual motion discrimination. Science (New N.Y. York,), *218*(4573), 697–698.

Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research*, *27*(6), 953–965.

Banai, K., & Amitay, S. (2012). Stimulus uncertainty in auditory perceptual learning. *Vision Research*, *61*, 83–88.

Bao, M., Fast, E., Mesik, J., & Engel, S. (2013). Distinct mechanisms control contrast adaptation over different timescales. *Journal of Vision*, *13*(10): 14, 1–11, https://doi.org/10.1167/13.10.14. [PubMed] [Article]

Bower, J. D., & Andersen, G. J. (2012). Aging, perceptual learning, and changes in efficiency of motion processing. *Vision Research*, *61*, 144–156.

Bi, J., Lee, H.-S., & O'Mahony, M. (2010). *d′* and variance of *d′* for four-alternative forced choice (4-AFC). *Journal of Sensory Studies*, *25*(5), 740–750.

Bi, T., Chen, J., Zhou, T., He, Y., & Fang, F. (2014). Function and structure of human left fusiform cortex are closely associated with perceptual learning of faces. *Current Biology*, *24*(2), 222–227.

Binns, A. M., Taylor, D. J., Edwards, L. A., & Crabb, D. P. (2018). Determining optimal test parameters for assessing dark adaptation in people with intermediate age-related macular degeneration. *Investigative Ophthalmology & Visual Science*, *59*(4), AMD114–AMD121.

Blair, C. A. J., & Hall, G. (2003). Perceptual learning in flavor aversion: Evidence for learned changes in stimulus effectiveness. *Journal of Experimental Psychology: Animal Behavior Processes*, *29*(1), 39–48.

Bower, J. D., Watanabe, T., & Andersen, G. J. (2013). Perceptual Learning and Aging: Improved Performance for Low-Contrast Motion Discrimination. *Frontiers in Psychology* , 4.

Camilleri, R., Pavan, A., Ghin, F., & Campana, G. (2014). Improving myopia via perceptual learning: is training with lateral masking the only (or the most) efficacious technique? *Attention, Perception & Psychophysics*, *76*(8), 2485–2494.

Cornsweet, T. N. (1962). The staircase method in psychophysics. *The American Journal of Psychology*, *75*(3), 485–491.

Donovan, I., Szpiro, S., & Carrasco, M. (2015). Exogenous attention facilitates location transfer of perceptual learning. *Journal of Vision*, *15*(10), 11.

Dosher, B. A., & Lu, Z.-L. (2007). The functional form of performance improvements in perceptual learning: Learning rates and transfer. *Psychological Science*, *18*(6), 531–539.

Dosher, B., & Lu, Z.-L. (2017). Visual Perceptual Learning and Models. *Annual Review of Vision Science*, *3*, 343–363.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, *70*(3), 193–242.

Fahle, M., & Edelman, S. (1993). Long-term learning in vernier acuity: Effects of stimulus orientation, range and of feedback. *Vision Research*, *33*(3), 397–412.

Fahle, M., Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision Research*, *35*(21), 3003–3013.

Fahle, M., & Morgan, M. (1996). No transfer of perceptual learning between similar stimuli in the same retinal position. *Current Biology: CB*, *6*(3), 292–297.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*(1), 585–612.

Greenland, S., & Kenneth, R. (1997). *Modern epidemiology* (2nd ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Hall, J. L. (1983). A procedure for detecting variability of psychophysical thresholds. *The Journal of the Acoustical Society of America*, *73*(2), 663–667.

Harris, H., Israeli, D., Minshew, N., Bonneh, Y., Heeger, D. J., Behrmann, M., & Sagi, D. (2015). Perceptual learning in autism: Over-specificity and possible remedies. *Nature Neuroscience*, *18*(11), 1574–1576.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponen-

tial law of practice. *Psychonomic Bulletin & Review*, *7*(2), 185–207.

Hou, F., Lesmes, L., Bex, P., Dorr, M., & Lu, Z.-L. (2015). Using 10AFC to further improve the efficiency of the quick CSF method. *Journal of Vision*, *15*(9):2, 1–18, https://doi.org/10.1167/15.9.2. [PubMed] [Article]

Huang, C.-B., Zhou, Y., & Lu, Z.-L. (2008). Broad bandwidth of perceptual learning in the visual system of adults with anisometropic amblyopia. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(10), 4068–4073.

Huang, J., Liang, J., Zhou, Y., & Liu, Z. (2017). Transfer in motion discrimination learning was no greater in double training than in single training. *Journal of Vision*, *17*(6):7, 1–10, https://doi.org/10.1167/17.6.7. [PubMed] [Article]

Hung, S.-C., & Seitz, A. R. (2014). Prolonged Training at Threshold Promotes Robust Retinotopic Specificity in Perceptual Learning. *Journal of Neuroscience*, *34*(25), 8423–8431.

Jackson, G. R., Owsley, C., & McGwin, G. (1999). Aging and dark adaptation. *Vision Research*, *39*(23), 3975–3982.

Jeter, P. E., Dosher, B. A., Liu, S.-H., & Lu, Z.-L. (2010). Specificity of perceptual learning increases with increased training. *Vision Research*, *50*(19), 1928–1940.

Jeter, P. E., Dosher, B. A., Petrov, A., & Lu, Z.-L. (2009). Task precision at transfer determines specificity of perceptual learning. *Journal of Vision*, *9*(3):1, 1–13, https://doi.org/10.1167/9.3.1. [PubMed] [Article]

Jogan, M., & Stocker, A. A. (2014). A new two-alternative forced choice method for the unbiased characterization of perceptual bias and discriminability. *Journal of Vision*, *14*(3):20, 1–18, https://doi.org/10.1167/14.3.20. [PubMed] [Article]

Kattner, F., Cochrane, A., & Green, C. S. (2017). Trial-dependent psychometric functions accounting for perceptual learning in 2-AFC discrimination tasks. *Journal of Vision*, *17*(11):3, 1–16, https://doi.org/10.1167/17.11.3. [PubMed] [Article]

Kingdom, F. A. A., & Prins, N. (2010). *Psychophysics: A practical introduction*. London: Elsevier.

King-Smith, P. E., & Rose, D. (1997). Principles of an adaptive method for measuring the slope of the psychometric function. *Vision Research*, *37*(12), 1595–1604.

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, *39*(16), 2729–2737.

Kujala, J. V., & Lukka, T. J. (2006). Bayesian adaptive estimation: The next dimension. *Journal of Mathematical Psychology*, *50*(4), 369–389.

Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics*, *63*(8), 1279–1292.

Lesmes, L. A., Jeon, S.-T., Lu, Z.-L., & Dosher, B. A. (2006). Bayesian adaptive estimation of threshold versus contrast external noise functions: The quick TvC method. *Vision Research*, *46*(19), 3160–3176.

Lesmes, L. A., Lu, Z.-L., Baek, J., & Albright, T. D. (2010). Bayesian adaptive estimation of the contrast sensitivity function: The quick CSF method. *Journal of Vision*, *10*(3):17, 1–21, https://doi.org/10.1167/10.3.17. [PubMed] [Article]

Lesmes, L. A., Lu, Z.-L., Baek, J., Tran, N., Dosher, B. A., & Albright, T. D. (2015). Developing Bayesian adaptive methods for estimating sensitivity thresholds (d') in yes-no and forced-choice tasks. *Frontiers in Psychology*, *6*, 1070.

Liang, J., Zhou, Y., Fahle, M., & Liu, Z. (2015). Limited transfer of long-term motion perceptual learning with double training. *Journal of Vision*, *15*(10):1, 1–9, https://doi.org/10.1167/15.10.1. [PubMed] [Article]

Liu, J., Lu, Z.-L., & Dosher, B. A. (2010). Augmented Hebbian reweighting: Interactions between feedback and training accuracy in perceptual learning. *Journal of Vision*, *10*(10), 29.

Liu, J., Lu, Z.-L., & Dosher, B. A. (2012). Mixed training at high and low accuracy levels leads to perceptual learning without feedback. *Vision Research*, *61*, 15–24.

Liu, Z. (1999). Perceptual learning in motion discrimination that generalizes across motion directions. *Proceedings of the National Academy of Sciences, USA*, *96*(24), 14085–14087.

Liu, Z. (1995). *Stimulus specificity in perceptual learning: Is it a consequence of experiments that are also stimulus specific?* Princeton, NJ: NEC Research Institute.

Liu, Z., & Weinshall, D. (2000). Mechanisms of generalization in perceptual learning. *Vision Research*, *40*(1), 97–109.

Lu, Z.-L., & Dosher, B. A. (2004). Perceptual learning retunes the perceptual template in foveal orientation identification. *Journal of Vision*, *4*(1), 44–56.

Lu, Z.-L., Chu, W., & Dosher, B. A. (2006). Perceptual learning of motion direction discrimination in fovea: Separable mechanisms. *Vision Research*, *46*(15), 2315–2327.

Lu, Z.-L., & Dosher, B. (2013). *Visual psychophysics:*

From laboratory to theory. Cambridge, MA: MIT Press.

Lu, Z.-L., Hua, T., Huang, C.-B., Zhou, Y., & Dosher, B. A. (2011). Visual perceptual learning. *Neurobiology of Learning and Memory*, *95*(2), 145–151.

Mackintosh, N. J., Kaye, H., & Bennett, C. H. (1991). Perceptual learning in flavour aversion conditioning. *Quarterly Journal of Experimental Psychology, B: Comparative and Physiological Psychology*, *43*(3), 297–322.

Mazur, J. E., & Hastie, R. (1978). Learning as accumulation: A reexamination of the learning curve. *Psychological Bulletin*, *85*(6), 1256–1274.

Meese, T. S. (1995). Using the standard staircase to measure the point of subjective equality: A guide based on computer simulations. *Perception & Psychophysics*, *57*(3), 267–281.

Mukai, I., Kim, D., Fukunaga, M., Japee, S., Marrett, S., & Ungerleider, L. G. (2007). Activations in Visual and Attention-Related Areas Predict and Correlate with the Degree of Perceptual Learning. *Journal of Neuroscience*, *27*(42), 11401–11411.

Nelles, G., Jentzen, W., Jueptner, M., Müller, S., & Diener, H. C. (2001). Arm training induced brain plasticity in stroke studied with serial positron emission tomography. *NeuroImage*, *13*(6 Pt 1), 1146–1154.

Monsen, E. R., & Horn, L. V. (2007). *Research: Successful approaches*. Chicago, IL: American Dietetic Association.

Moore, D. R., Amitay, S., & Hawkey, D. J. C. (2003). Auditory perceptual learning. *Learning & Memory*, *10*(2), 83–85.

Perez, C., & Chokron, S. (2014). Rehabilitation of homonymous hemianopia: insight into blindsight. *Frontiers in Integrative Neuroscience*, 8.

Petrov, A. A., Dosher, B. A., & Lu, Z.-L. (2005). The dynamics of perceptual learning: An incremental reweighting model. *Psychological Review*, *112*(4), 715–743.

Polat, U., Ma-Naim, T., Belkin, M., & Sagi, D. (2004). Improving vision in adult amblyopia by perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(17), 6692–6697.

Rodríguez, G., & Angulo, R. (2014). Simultaneous stimulus preexposure enhances human tactile perceptual learning. *Psicológica*, *35*(1), 139–148.

Sagi, D. (2011). Perceptual learning in vision research. *Vision Research*, *51*(13), 1552–1566.

Sasaki, Y., Náñez, J. E., & Watanabe, T. (2010). Advances in visual perceptual learning and plasticity. *Nature Reviews Neuroscience*, *11*(1), 53–60.

Sasaki, Y., Náñez, J. E., & Watanabe, T. (2012). Recent progress in perceptual learning research. *WIREs Cognitive Science*, *3*(3), 293–299.

Sathian, K., & Zangaladze, A. (1998). Perceptual learning in tactile hyperacuity: Complete intermanual transfer but limited retention. *Experimental Brain Research*, *118*(1), 131–134.

Scahill, V. L., & Mackintosh, N. J. (2004). The easy to hard effect and perceptual learning in flavor aversion conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *30*(2), 96–103.

Schoups, A. A., Vogels, R., & Orban, G. A. (1995). Human perceptual learning in identifying the oblique orientation: Retinotopy, orientation specificity and monocularity. *The Journal of Physiology*, *483*(3), 797–810.

Seitz, A. R., Kim, D., & Watanabe, T. (2009). Rewards evoke learning of unconsciously processed visual stimuli in adult humans. *Neuron*, *61*(5), 700–707.

Shelton, B. R., & Scarrow, I. (1984). Two-alternative versus three-alternative procedures for threshold estimation. *Perception & Psychophysics*, *35*(4), 385–392.

Shibata, K., Yamagishi, N., Ishii, S., & Kawato, M. (2009). Boosting perceptual learning by fake feedback. *Vision Research*, *49*(21), 2574–2585.

Sowden, P. T., Rose, D., & Davies, I. R. L. (2002). Perceptual learning of luminance contrast detection: Specific for spatial frequency and retinal location but not orientation. *Vision Research*, *42*(10), 1249–1258.

Stevenson, R. J. (2001). Associative learning and odor quality perception: How sniffing an odor mixture can alter the smell of its parts. *Learning and Motivation*, *32*(2), 154–177.

Symonds, M., & Hall, G. (1995). Perceptual learning in flavor aversion conditioning: Roles of stimulus comparison and latent inhibition of common stimulus elements. *Learning and Motivation*, *26*(2), 203–219.

Szpiro, S. F. A., & Carrasco, M. (2015). Exogenous Attention Enables Perceptual Learning. *Psychological Science*, *26*(12), 1854–1862.

Tan, D. T. H., & Fong, A. (2008). Efficacy of neural vision therapy to enhance contrast sensitivity function and visual acuity in low myopia. *Journal of Cataract and Refractive Surgery*, *34*(4), 570–577.

Tartaglia, E. M., Bamert, L., Mast, F. W., & Herzog, M. H. (2009). Human perceptual learning by

mental imagery. *Current Biology*, *19*(24), 2081–2085.

Thurston, C., & Dobkins, K. (2007). Stimulus-specific perceptual learning for chromatic, but not luminance, contrast detection. *Journal of Vision*, *7*(9): 469, https://doi.org/10.1167/7.9.469. [Abstract]

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

Vancleef, K., Read, J. C. A., Herbert, W., Goodship, N., Woodhouse, M., & Serrano-Pedraza, I. (2018). Two choices good, four choices better: For measuring stereoacuity in children, a four-alternative forced-choice paradigm is more efficient than two. *PLoS One*, *13*(7): e0201366.

Wang, R., Wang, J., Zhang, J.-Y., Xie, X.-Y., Yang, Y.-X., Luo, S.-H., … Li, W. (2016). Perceptual learning at a conceptual level. *The Journal of Neuroscience*, *36*(7), 2238–2246.

Ward, L. M., Morison, G., Simmers, A. J., & Shahani, U. (2018). Age-related changes in global motion coherence: Conflicting haemodynamic and perceptual responses. *Scientific Reports*, *8*(1), 10013.

Watanabe, T., & Sasaki, Y. (2015). Perceptual learning: Toward a comprehensive theory. *Annual Review of Psychology*, *66*, 197–221.

Watson, A. B. (2017). QUEST+: A general multidimensional Bayesian adaptive psychometric method. *Journal of Vision*, *17*(3):10, 1–27, https://doi.org/10.1167/17.3.10. [PubMed] [Article]

Watson, A. B., & Pelli, D. G. (1979). The QUEST staircase procedure. *Applied Vision Association Newsletter*, *14*, 6–7.

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

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

Wilson, D. A., & Stevenson, R. J. (2003). The fundamental role of memory in olfactory perception. *Trends in Neurosciences*, *26*(5), 243–247.

Wright, B. A., & Zhang, Y. (2009). A review of the generalization of auditory learning. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1515), 301–311.

Xiao, L.-Q., Zhang, J.-Y., Wang, R., Klein, S. A., Levi, D. M., & Yu, C. (2008). Complete transfer of perceptual learning across retinal locations enabled by double training. *Current Biology*, *18*(24), 1922–1926.

Yan, F.-F., Zhou, J., Zhao, W., Li, M., Xi, J., Lu, Z.-L., & Huang, C.-B. (2015). Perceptual learning improves neural processing in myopic vision. *Journal of Vision*, *15*(10).

Zhang, P., Hou, F., Yan, F.-F., Xi, J., Lin, B.-R., Zhao, J., … Huang, C.-B. (2018). High reward enhances perceptual learning. *Journal of Vision*, *18*(8):11, 1–21, https://doi.org/10.1167/18.8.11. [PubMed] [Article]

Zhang, P., Zhao, Y., Dosher, B., & Lu, Z.-L. (2018). Assessing the trial-by-trial time course of perceptual sensitivity change in perceptual learning using the quick Change Detection method. *Journal of Vision*, *18*(10): 1068, https://doi.org/10.1167/18.10.1068. [Abstract]

Zhang, P., Zhao, Y., Dosher, B. A., & Lu, Z.-L. (2019). Assessing the detailed time course of perceptual sensitivity change in perceptual learning. *Journal of Vision*, *19*(5):9, 1–19, https://doi.org/10.1167/19.5.9. [PubMed] [Article]

Zhao, Y., Lesmes, L. A., & Lu, Z.-L. (2017). The quick Change Detection method: Bayesian adaptive assessment of the time course of perceptual sensitivity change. *Investigative Ophthalmology & Visual Science*, *58*(8), 5633–5633.

Zhao, Y., Lesmes, L. A., & Lu, Z.-L. (2019). Efficient assessment of the time course of perceptual sensitivity change. *Vision Research*, *154*, 21–43.

Zhou, Y., Huang, C., Xu, P., Tao, L., Qiu, Z., Li, X., & Lu, Z.-L. (2006). Perceptual learning improves contrast sensitivity and visual acuity in adults with anisometropic amblyopia. *Vision Research*, *46*(5), 739–750.

Appendix A: The qCD procedure

Step 1. Threshold, psychometric function, and priors

In the qCD method, the true learning curve is characterized with a single exponential function:

$$T\left(\vec{\theta}, n\right) = \lambda exp\left(\frac{-n}{\gamma}\right) + \alpha, \quad \text{(A1)}$$

where $\vec{\theta} = (\theta_1, \theta_2, \theta_3) = (\lambda, \gamma, \alpha)$, n is the trial number during training, $T(\vec{\theta}, n)$ is the threshold at the $d' = 1.5$ performance level on trial n, λ is the dynamic range of learning, α is the asymptotic threshold level, and γ is the time constant of the exponential function. Therefore, $\lambda + \alpha$ is the initial threshold before training. A broad joint prior distribution $p_0(\vec{\theta})$ is defined in the three-

dimensional parameter space of $\vec{\theta}$, where the subscript 0 is used to denote that the prior that represents a priori knowledge of the parameters of the learning curve before the experiment.

The prior distribution $p_0(\vec{\theta})$ was defined by the hyperbolic secant function

$$p_0\left(\vec{\theta}\right) = \operatorname{sech}\left(\lambda_{confidence}(log_{10}(\lambda) - log_{10}(\lambda_{mode}))\right) \times \operatorname{sech}\left(\gamma_{confidence}(log_{10}(\gamma) - log_{10}(\gamma_{mode}))\right) \times \operatorname{sech}\left(\alpha_{confidence}(log_{10}(\alpha) - log_{10}(\alpha_{mode}))\right) \quad \text{(A2)}$$

where $\operatorname{sech}(x) = 2/(e^x + e^{-x})$; λ_{mode}, γ_{mode}, and α_{mode} are the peaks of the respective secant functions; and $\lambda_{confidence}$, $\gamma_{confidence}$, and $\alpha_{confidence}$ are the spreads of the respective secant functions. The secant function is used as a weakly informative prior based on the previous literature (King-Smith & Rose, 1997) and results from a pilot study. The joint prior distribution $p_0(\vec{\theta})$ was updated trial by trial throughout the simulated experiment. A one-dimensional space X covers all possible levels of the stimulus variation $x(\in X)$.

In a two-alternative forced-choice task, the percentage-correct psychometric function is described by a Weibull function (Wichmann & Hill, 2001):

$$p'_n\left(r = 1|\vec{\theta}, x\right) = g + (1 - g)\left(1 - \exp\left(-\left(\frac{x}{T_w\left(\vec{\theta}, n\right)}\right)^s\right)\right), \quad \text{(A3)}$$

$$p_n\left(r = 1|\vec{\theta}, x\right) = (1 - \lambda)p'_n\left(r = 1|\vec{\theta}, x\right) + lg, \quad \text{(A4)}$$

where p'_n is the psychometric function on trial n without a lapse rate, and p_n is with a lapse rate, r is the response (1 for correct, 0 for incorrect), x is the stimulus value, and $g = 0.5$ is the guessing rate. The slope of the Weibull function was set as $s = 3.06$ (Hou, Lesmes, Bex, Dorr, & Lu, 2015), and the lapse rate was set at $l = 0.04$. In this case, $T_w(\theta, n)$ is the Weibull threshold at trial n given parameters $\vec{\theta}$:

$$log_{10}\left(T_w\left(\vec{\theta}, n\right)\right) = log_{10}\left(T\left(\vec{\theta}, n\right)\right) - \frac{1}{s}log_{10}\left(\log\left(\frac{1 - g}{1 - p_{1.5}}\right)\right), \quad \text{(A5)}$$

where $p_{1.5} = 0.794$ is the probability correct corresponding to $d' = 1.5$ in a two-alternative forced-choice task. Therefore, $p_n(r = 1|\vec{\theta}, x)$ is the probability of making a correct response ($r = 1$) at trial n, conditioned on parameters $\vec{\theta}$ and the stimulus value x. The probability of making an incorrect response ($r = 0$) is

$$p_n\left(r = 0|\vec{\theta}, x\right) = 1 - p_n\left(r = 1|\vec{\theta}, x\right). \quad \text{(A6)}$$

Together, Equations A1 through A6 describe the probability that any possible observer (defined by $\vec{\theta}$) will make either a correct or an incorrect response in all possible stimulus conditions (x) throughout the course of perceptual learning. The qCD method then uses an adaptive procedure to update the posterior distributions for a given observer based on their performance in the simulation.

Step 2. One-step-ahead stimulus search

The stimulus in the next trial is selected from among all possible stimulus values (the one-dimensional space X) to optimize the expected information gain (Kujala & Lukka, 2006; Lesmes, Jeon, Lu, & Dosher, 2006). The expected information gain of a potential test stimulus with x is defined as

$$I_n(x) = h\left(\sum_{\vec{\theta}} p_n\left(\vec{\theta}\right)p_n\left(r = 1|\vec{\theta}, x\right)\right) - \sum_{\vec{\theta}} p_n\left(\vec{\theta}\right)h\left(p_n\left(r = 1|\vec{\theta}, x\right)\right), \quad \text{(A7)}$$

$$h(p) = -p\log(p) - (1 - p)\log(1 - p), \quad \text{(A8)}$$

where h is the information entropy of the distribution p.

Step 3. Bayesian update

After the stimulus for the next trial (e.g., trial n) is selected by Step 2 and presented to the observer, the joint posterior distribution of the parameters of the exponential function is updated based on the response of the observer.

In the nth trial of the learning curve, r_n is the observer's response to a stimulus with stimulus value x_n presented at trial n of training, and $p_n(\vec{\theta})$ is the prior distribution of parameters $\vec{\theta}$ at trial n of training (equal to the posterior for trial $n - 1$). The prior distribution $p_n(\vec{\theta})$ is updated to the posterior distribution $p_n(\vec{\theta}|r_n, x_n)$ via Bayes's rule:

$$p_n\left(\vec{\theta}|r_n, x_n\right) = \frac{p_n(r_n|\vec{\theta}, x_n)\ p_n\left(\vec{\theta}\right)}{p_n(r_n|x_n)}, \quad \text{(A9)}$$

$$p_n(r_n|x_n) = \sum_{\vec{\theta}} p_n(r_n|\vec{\theta}, x_n)\ p_n\left(\vec{\theta}\right), \quad \text{(A10)}$$

where $\Sigma_{\vec{\theta}}$ is the summation over the three-dimensional parameter space—that is, $\Sigma_{\vec{\theta}} = \Sigma_\alpha\Sigma_\lambda\Sigma_\gamma$.

The posterior distribution of $\vec{\theta}$ following the nth trial in the learning curve is used as the prior of trial $n+1$:

$$p_{n+1}\left(\vec{\theta}\right) = p_n\left(\vec{\theta}|r_n, x_n\right). \quad \text{(A11)}$$

The marginal posterior distribution of each parameter is computed via summation of the joint posterior distribution over the other two parameters in the three-dimensional parameter space:

$$p_n(\alpha|r_n, x_n) = \sum_{\lambda}\sum_{\gamma} p_n\left(\vec{\theta}|r_n, x_n\right)$$

$$p_n(\lambda|r_n, x_n) = \sum_{\alpha}\sum_{\gamma} p_n\left(\vec{\theta}|r_n, x_n\right) \quad \text{(A12)}$$

$$p_n(\gamma|r_n, x_n) = \sum_{\alpha}\sum_{\lambda} p_n\left(\vec{\theta}|r_n, x_n\right).$$

The expected values of the marginal posterior distributions provide estimates of the parameters of the exponential function after the nth trial in the learning curve:

$$\bar{\theta}_{f,n} = \sum_{\theta_f} \theta_f \cdot p_n(\theta_\alpha|r_n, x_n), \quad \text{(A13)}$$

where $\theta_f = \lambda, \gamma, \alpha$ for $f = 1, 2, 3$, respectively. (These are cast as summations rather than integrals because the parameter values for which these quantities are computed are quantized in the qCD method and in the simulations; see main text.)

Based on the updated joint posterior distribution of parameters, the trial threshold can be estimated by the following procedures. We constructed 1,000 learning curves from the 1,000 sets of parameters randomly sampled based on the posterior probability distributions of the parameters. The average of the 1,000 learning curves was used as the threshold estimates.

Step 4. Iteration of Steps 2 and 3

Steps 2 and 3 are iterated until a predefined number of training trials are completed.

Step 5. Trial-by-trial and segment-by-segment analyses

Based on the joint posterior distribution after each trial, qCD can provide trial-by-trial estimates of the parameters and a single threshold in that trial. In addition, a post hoc segment-by-segment analysis can be used to partition the posterior distributions into segments based on their central tendency. The post hoc segment-by-segment analysis can determine whether the entire learning curve can be described by a single exponential function or a cascade of multiple exponent functions, and use all the information collected during each segment to further improve the accuracy and precision of the estimated thresholds (Zhao, Lesmes, & Lu, 2017, 2019). The posterior distribution obtained at the end of each segment is used to compute estimates of parameters and thresholds.

We can partition the posterior distributions based on their central tendency across time. The distance between the central tendencies of the posterior distributions $p_{n_1}(\vec{\theta}|r_{n_1}, x_{n_1})$ and $p_{n_2}(\vec{\theta}|r_{n_2}, x_{n_2})$ at two trials, n_1 and n_2, is quantified by a modified Mahalanobis distance:

$$MD(n_1, n_2) = \sqrt{\vec{y}C^{-1}\vec{y}^{tr}}, \quad \text{(A14)}$$

$$C = [C_{n_1} + C_{n_2}]/2, \quad \text{(A15)}$$

where $\vec{y} = [y_1, y_2, y_3]$, $y_i = \bar{\theta}_{i,n_1} - \bar{\theta}_{i,n_2}$, $\theta_f = \lambda, \gamma, \alpha$,for $f =$ 1, 2, 3, respectively; $\vec{y}^T$ is the transpose of $\vec{y}$; C^{-1} is the inverse of the matrix C; and C_{n_b} ($b = 1$ or 2) is the 3×3 covariance matrix of the posterior $p_{n_b}(\vec{\theta}|r_{n_b}, x_{n_b})$ for which the diagonal elements c_{n_b,a_1a_1} and the off-diagonal elements $c_{n_b,ij}$ ($i, j \in [1, 2, 3], i \neq j$) are defined as

$$\begin{aligned} c_{n_b,ii} &= Var\left(\theta_{i,n_b}\right) \\ &= \sum_{\vec{\theta}} \left(\theta_i - \bar{\theta}_{i,n_b}\right)^2 \cdot p_{n_b}\left(\vec{\theta}|r_{n_b}, x_{n_b}\right), \end{aligned} \quad \text{(A16)}$$

$$\begin{aligned} c_{n_b,ij} &= Cov\left(\theta_{i,n_b}, \theta_{j,n_b}\right) \\ &= \sum_{\vec{\theta}} \left(\theta_i - \bar{\theta}_{i,n_b}\right)\left(\theta_j - \bar{\theta}_{j,n_b}\right) \\ &\quad \cdot p_{n_b}\left(\vec{\theta}|r_{n_b}, x_{n_b}\right). \end{aligned} \quad \text{(A17)}$$

The null hypothesis is that the posterior distributions $p_{n_1}(\vec{\theta}|r_{n_1}, x_{n_1})$ and $p_{n_2}(\vec{\theta}|r_{n_2}, x_{n_2})$ are the same. We reject the null hypothesis when

$$MD(n_1, n_2) > MD_0, \quad \text{(A18)}$$

where MD_0 is a predetermined criterion.

To describe how we partition the posterior distributions of the dark-adaptation curve, we first introduce some notation: u_l is the trial number in the experiment, where l and u_l refer to the lth segment and the u_lth trial in the lth segment, respectively; U_l is the total number of trials in the lth segment; and L is the number of segments of the entire dark-adaptation curve. To partition the dark-adaptation curve into segments, we start from the last trial of the entire experiment, U_L.

The MD between $p_{n_{U_L}}(\vec{\theta}|r_{U_L}, x_{U_L})$ and $p_{n_{u_L}}(\vec{\theta}|r_{u_L}, x_{u_L})$, the posterior distributions of the last trial and the previous trials, is calculated until U_{L-1}, the last trial of segment $L-1$, is found:

$$MD\left(p_{n_{U_L}}\left(\vec{\theta}|r_{U_L}, x_{U_L}\right), p_{n_{u_L}}\left(\vec{\theta}|r_{u_L}, x_{u_L}\right)\right) \leq MD_0, \quad \forall u_L \in [1, U_L - 1], \quad \text{(A19)}$$

$$MD\left(p_{n_{U_L}}\left(\vec{\theta}|r_{U_L}, x_{U_L}\right), p_{n_{U_{L-1}}}\left(\vec{\theta}|r_{U_{L-1}}, x_{U_{L-1}}\right)\right) > MD_0, \quad \text{(A20)}$$

where n_{u_L} and n_{U_L} refer to trials u_L and U_L (last trial) in the Lth segment (last segment).

We then repeat the procedure to find all the segments. Therefore, for the lth segment ($l > 1$), we have

$$MD\left(p_{n_{U_l}}\left(\vec{\theta}|r_{U_l}, x_{U_l}\right), p_{n_{u_l}}\left(\vec{\theta}|r_{u_l}, x_{u_l}\right)\right) \leq MD_0, \quad \forall u_l \in [1, U_l - 1], \quad \text{(A21)}$$

$$MD\left(p_{n_{U_l}}\left(\vec{\theta}|r_{U_l}, x_{U_l}\right), p_{n_{U_{l-1}}}\left(\vec{\theta}|r_{U_{l-1}}, x_{U_{l-1}}\right)\right) > MD_0. \quad \text{(A22)}$$

Because the learning curve is continuous, we impose the following constraint on the posterior distributions between segments:

$$p_{u_l}(\lambda + \alpha) = p_{n_{U_{l-1}}}\left(f_{t_{U_{l-1}}}\right), \quad \text{(A23)}$$

where $p_{u_l}(\lambda + \alpha)$ is the marginal prior distribution of $\lambda + \alpha$ at trial u_l (u_lth trial in the lth segment) and $p_{n_{U_{l-1}}}$ ($f_{n_{U_{l-1}}}$) is the probability distribution of the threshold estimate at $n_{\{U_{l-1}, l-1\}}$ of segment $l-1$.

Following segmentation, we use the posterior in the last trial of each segment, $p_{n_{U_l}}(\vec{\theta}|r_{U_l}, x_{U_l})$, to compute the estimated thresholds in the entire segment. Therefore, the posteriors are conditional on the data from the beginning of the experiment.