Assessing the detailed time course of perceptual sensitivity change in perceptual learning

Pan Zhang

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

Yukai Zhao

Laboratory of Brain Processes (LOBES), Departments 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), Departments of Psychology, The Ohio State University, Columbus, OH, USA ✉

The learning curve in perceptual learning is typically sampled in blocks of trials, which could result in imprecise and possibly biased estimates, especially when learning is rapid. Recently, Zhao, Lesmes, and Lu (2017, 2019) developed a Bayesian adaptive quick Change Detection (qCD) method to accurately, precisely, and efficiently assess the time course of perceptual sensitivity change. In this study, we implemented and tested the qCD method in assessing the learning curve in a four-alternative forced-choice global motion direction identification task in both simulations and a psychophysical experiment. The stimulus intensity in each trial was determined by the qCD, staircase or random stimulus selection (RSS) methods. Simulations showed that the accuracy (bias) and precision (standard deviation or confidence bounds) of the estimated learning curves from the qCD were much better than those obtained by the staircase and RSS method; this is true for both trial-by-trial and post hoc segment-by-segment qCD analyses. In the psychophysical experiment, the average half widths of the 68.2% credible interval of the estimated thresholds from the trial-by-trial and post hoc segment-by-segment qCD analyses were both quite small. Additionally, the overall estimates from the qCD and staircase methods matched extremely well in this task where the behavioral rate of learning is relatively slow. Our results suggest that the qCD method can precisely and accurately assess the trial-by-trial time course of perceptual learning.

Introduction

Research in the last 30 years has revealed a great deal about the characteristics and mechanisms of perceptual learning (Dosher & Lu, 2017; Green, Banai, Lu, & Bavelier, 2018; Sagi, 2011; Sasaki, Nanez, & Watanabe, 2010; Seitz, 2017). It has been shown that perceptual learning can improve human performance over a wide range of perceptual tasks, with relatively long-lasting changes to the perceptual system (Goldstone, 1998; Zhou et al., 2006). The benefits from training or practice are often specific or partially specific to the characteristics of the trained stimulus or task (Fiorentini & Berardi, 1980; Karni & Sagi, 1991), with the degree of transfer or specificity depending critically on the difficulty (Ahissar & Hochstein, 1997; Z. Liu, 1995, 1999) or precision (Jeter, Dosher, Petrov, & Lu, 2009) of the training and/or transfer tasks, and training protocol (J. Huang, Liang, Zhou, & Liu, 2017; Hung & Seitz, 2014; Liang, Zhou, Fahle, & Liu, 2015; Xiao et al., 2008). In the meantime, a large number of studies in psychophysics (Dosher & Lu, 1998, 1999; Gold, Bennett, & Sekuler, 1999; Saarinen & Levi, 1995) and neurophysiology (Crist, Li, & Gilbert, 2001; Ghose, Yang, & Maunsell, 2002; Schoups, Vogels, Qian, & Orban, 2001) have focused on the mechanism of perceptual learning. There is also an active market place for perceptual learning (Lu, Lin, & Dosher,

2016). Overall, these findings of perceptual learning have revealed important plasticity of the adult human perceptual system and become an integral component of our understanding of perception.

One fundamental building block in all the perceptual learning studies is the learning curve—that is, how perceptual sensitivity changes as a function of training. The learning curve is important not only for estimating the magnitude of learning and the degree of specificity and transfer (Ahissar & Hochstein, 1997; C.-B. Huang, Lu, & Dosher, 2012; Jeter et al., 2009; Z. Liu & Weinshall, 2000), but also for specifying the functional form of learning (Dosher & Lu, 2007; Poggio, Fahle, & Edelman, 1992). Comparisons of learning curves obtained under different external noise (Dosher & Lu, 2005; Lu, Chu, & Dosher, 2006), training difficulty (J. Liu, Lu, & Dosher, 2010; J. Liu, Lu, & Dosher, 2012; Z. Liu & Weinshall, 2000), training schedule (Hung & Seitz, 2014; Xiao et al., 2008), feedback (Fahle & Edelman, 1993; Herzog & Fahle, 1997; J. Liu et al., 2010; J. Liu et al., 2012; Shibata, Yamagishi, Ishii, & Kawato, 2009), attention (Donovan, Szpiro, & Carrasco, 2015; Mukai et al., 2007; Szpiro & Carrasco, 2015; Szpiro, Lee, Wright, & Carrasco, 2013), and reward conditions (Seitz, Kim, & Watanabe, 2009; Zhang, Hou, et al., 2018) can reveal the impact of the manipulations on perceptual learning and inform and constrain computational models of perceptual learning (Dosher & Lu, 2017; Lu, Liu, & Dosher, 2010; Petrov, Dosher, & Lu, 2005). Unbiased, precise, and detailed assessment of the learning curve is critical in perceptual learning research.

Three types of performance measures—accuracy (percent correct or *d*-primes), contrast thresholds, and difference thresholds (Ball, Sekuler, & Machamer, 1983; C.-B. Huang, Zhou, & Lu, 2008; Karni & Sagi, 1991; Leek, 2001; Levi & Polat, 1996; Z. Liu, 1999; Pelli & Bex, 2013; Pelli & Farell, 1995; Petrov et al., 2005)—are often used to construct learning curves. Existing methods for assessing all three performance measures are based on blocks of measurements with relatively large numbers of trials. The basic assumption of many of these methods is that performance does not change within each measurement block and that some form of averaging can be used to gauge the performance in the block. However, because performance may change continuously during perceptual learning (Lu, Hua, Huang, Zhou, & Dosher, 2011; Mazur & Hastie, 1978; Petrov et al., 2005), even within each measurement block, especially in the early phase of learning, the resulting learning curves can be imprecise and the measurements may be biased; this in turn may, in some circumstances, lead to incorrect inferences about properties of perceptual learning.

Take as an example the up-down staircase method, originally introduced as a method to estimate a fixed threshold (Leek, 2001; Levitt, 1971). In perceptual learning studies, the up-down staircase method is often used to estimate the threshold at a certain percent correct performance level by adjusting the stimulus level (e.g., contrast, luminance, orientation difference, motion coherence) based on the observer's responses (Ahissar & Hochstein, 1996, 1997; Dosher & Lu, 1998; Polat, Ma-Naim, Belkin, & Sagi, 2004; Xiao et al., 2008). Specifically, the method begins with an initial stimulus level, usually chosen based on prior knowledge about the observer and task, then increases or decreases the stimulus level with either fixed or adaptive step sizes based on the observer's response in each trial, and stops after a predefined number of trials or number of reversals. A single threshold is obtained by averaging the endpoints of multiple reversals of the staircase in each measurement block, and the learning curve is constructed from the estimated thresholds over blocks. There are three major problems associated with the staircase procedure as applied to situations where the threshold may be changing (Zhang, Zhao, Dosher, & Lu, 2018): (a) It often takes 60–80 trials for the staircase procedure to converge and therefore a measurement block usually contains 60–80 trials. This limits the temporal resolution of the estimated learning curve, which, in certain cases, may miss rapid learning in the early phase of training if it occurs (Fahle, Edelman, & Poggio, 1995; Hawkey, Amitay, & Moore, 2004; Poggio et al., 1992). (b) It may lead to biased estimates of the initial performance level, learning rate, and specificity/transfer index (Lu, Zhang, Zhao, & Dosher, 2018). (c) Choosing the optimal starting stimulus level and step size is very challenging (Lu & Dosher, 2013). Nonoptimal settings can increase the bias even in estimates of fixed thresholds.

The importance of a detailed estimate of the time course of perceptual learning has been previously argued in the literature (Kumar & Glaser, 1993). To obtain a detailed time course of learning, Kumar and Glaser (1993) averaged observers' trial-by-trial performance in hyperacuity tasks over 34 different nonstereo and 49 stereo hyperacuity stimuli. By using this trial-by-trial method, they discovered very rapid learning (within the first few trials) for nonstereo stimuli, and learning that was far slower for the stereo stimuli. This procedure, which averages over many stimuli and training runs, is, however, not practical in most perceptual learning studies because large numbers of different stimuli in the same task category are often not available.

In another approach to estimating a more detailed time course of the learning curve, Kattner, Cochrane, and Green (2017) fitted data obtained with the method of constant stimuli using psychometric functions with learning dependent parameters. In their fitting procedure, the threshold was modeled as a function of time/
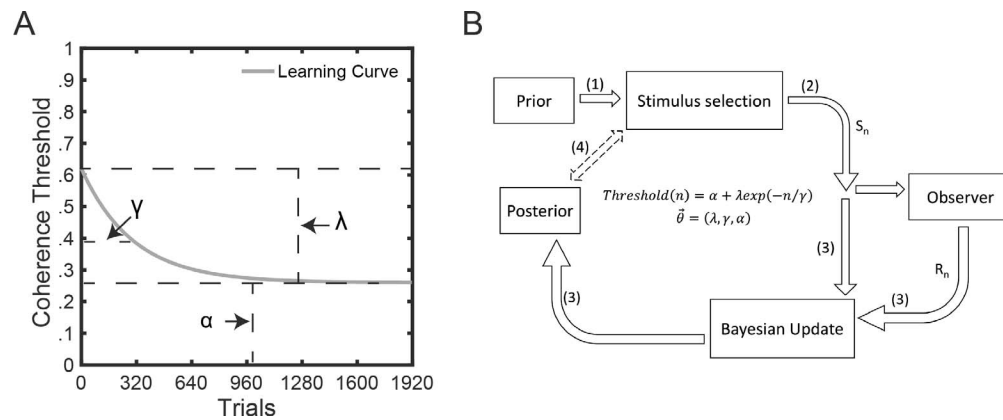
Figure 1. (A) Exponential function with three parameters: (1) λ: dynamic range of learning; (2) γ: time constant of exponential function; and (3) α: asymptotic performance level. (B) The qCD method consists of five steps: (1) The time course of motion coherence threshold is defined as an exponential function (as a function of trial number n) with three parameters $\vec{\theta} = (\lambda, \gamma, \alpha)$, and their joint prior distribution; (2) The proportion of coherent dots 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' rule based on the 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.

trial number. Although this 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, there are still limitations because the selection of test stimuli during learning was not optimized for rapid assessment. With the method of constant stimuli used in the study, the stimulus level in each trial was randomly sampled from a uniform distribution in a predetermined range (random stimulus selection, RSS). As learning proceeded, more and more of the stimuli resulted in ceiling level performance and so provided very little new information about, thus not much constraint on, the psychometric function. Although the new fitting procedure significantly improved the confidence intervals of the estimated thresholds in the late phase of the learning curve, it did not substantially benefit estimates in the early phase of learning.

Here, we implemented and validated a Bayesian adaptive procedure, the quick change detection (qCD, see Figure 1) method developed by Zhao, Lesmes, and Lu (2017, 2019) to measure the learning curve in perceptual learning. The qCD method was designed to accurately, precisely, and efficiently measure the time course of perceptual sensitivity change. Unlike existing procedures that assume that perceptual sensitivity does not change within each measurement block or epoch (Leek, 2001; Lesmes, Jeon, Lu, & Dosher, 2006; Lesmes, Lu, Baek, & Albright, 2010; Lesmes et al., 2015; Taylor & Creelman, 1967; Treutwein, 1995; Watson & Pelli, 1983), the qCD method explicitly models and estimates how perceptual sensitivity changes over time. Using the Bayesian adaptive testing framework (Lesmes et al., 2006; Lesmes et al., 2010;

Lesmes et al., 2015; Watson & Pelli, 1983), the method selects the optimal stimulus—the stimulus for which the response gives the most of new information—and updates, trial by trial, a joint probability distribution of the parameters of a model of perceptual sensitivity change over time. In its first implementation in dark adaptation, Zhao et al. (2017, 2019) demonstrated with computer simulations of an eight-alternative forced-choice (8AFC) task in which one run of the qCD method can be used to estimate the dark adaptation curve with better accuracy and precision than 10 runs of the quick Forced Choice (qFC; Lesmes et al., 2015) and staircase methods. Furthermore, the dark adaptation curve obtained from one qCD run in a psychophysics experiment was highly consistent with the average of four qFC runs (RMSE = 0.076 $\log_{10}$ units).

Zhao et al. (2017, 2019) developed the general qCD method and evaluated the method with computer simulations and a psychophysical task for dark adaptation in an 8AFC task. In the current paper, we further evaluated the qCD method in simulations and a psychophysical experiment in perceptual learning in a four-alternative forced-choice (4AFC) task. There are several important new aspects of the current study. First, unlike the dark adaptation curve, the learning curve in perceptual learning cannot be measured repeatedly. This makes the qCD method even more valuable, but also poses unique challenges for validating the method. In the current study, we developed an interleaved validation procedure in which we alternated qCD trials with staircase trials. Second, the performance of the qCD method depends on the number of alternatives in the perceptual task. Its performance in 4AFC must be independently evaluated. Third, we

	Observer 1	Observer 2	Observer 3
λ	0.385	0.385	0.385
γ	40	80	160
α	0.275	0.275	0.275

Table 1. Parameters of the three simulated observers.

compared the performance of the qCD method with that of a staircase procedure in a range of initial stimulus intensity levels.

Simulations

Method

To evaluate the performance of the qCD method in assessing the exponential learning curve (see the function Figure 1B), we simulated three observers with different learning parameters with the qCD, staircase, and RSS methods (Table 1). These simulations were set up to approximate performance in a global motion direction identification task used for experimental validation (Figure 2).

A description of the qCD method is provided in Supplementary Appendix A. The prior of the parameters of the exponential function in the qCD method in the simulation study was set up based on results from a pilot behavior experiment. The parameter space included 50 log-linearly spaced λ values (from 0.05 to 0.7), 50 log-linearly spaced γ values (from 20 to 600), and 50 log-linearly spaced α values (from 0.1 to 0.4). (The λ and α are in the units of the threshold measurement—here proportion of coherent dots— while the γ values are in units of trials. Also, in the behavioral experiment, the discrete values of proportions of coherent dots that can be programmed depends on the total number of dots). For λ, 0 was also included to account for no learning. $(\lambda_{mode}, \gamma_{mode}, \alpha_{mode}) = (0.36, 326, 0.26)$ are the modes of the respective secant functions; $(\lambda_{confidence}, \gamma_{confidence}, \alpha_{confidence}) = (5.81, 3.82, 12.67)$ are the spreads of the respective secant functions. A one-dimension space X included 100 log-linearly spaced values from 0.01 to 1.

The simulated observers performed a 4AFC task. One thousand simulated runs were conducted for each observer. Each run consisted of 960 trials of the qCD method and 960 trials of the three-down/one-up staircase method (Levitt, 1971), arranged in alternation. Therefore, the odd trials (e.g., 1, 3, 5, . . . and 1919) and even trials (e.g., 2, 4, 6, . . . and 1920) were simulated by the qCD and staircase methods, respectively. No information was shared between the qCD and staircase procedure, and they were run independently and separately. In each simulated trial, the *true* threshold, T(n), was calculated using Equation 1. Then, the expected probability of making a correct response was calculated for the stimulus level x using Equation 4 (e.g., given the value of the percent correct psychometric function on trial n). To determine if the observer's response is correct or not on that trial, we first drew a random number r from a uniform distribution over the interval from 0 to 1, and then labeled the response as correct if $r < p_n(r = 1|\vec{\theta}, x)$, and incorrect otherwise.

In the staircase procedure, if the simulated observer makes three consecutive correct responses (in the staircase trials), the proportion of coherent dots is reduced by 10% (multiplied by 0.9) in the next trial; if the simulated observer makes a single incorrect response, the proportion of coherent dots is increased by 10% (multiplied by 1.1) in the next trial. A reversal happens if the staircase changes its direction (from up to down or vice versa). Usually, one block of three-down/one-up staircase with a 10% step size produces about a dozen "reversals" (or endpoints) in 80 trials. The estimated threshold of the block is calculated by averaging the proportion of coherent dots at the reversals after deleting the first four or five of them.

To compare the performance of the qCD with that of the RSS method, we conducted a simulation study using the RSS method on the same three simulated
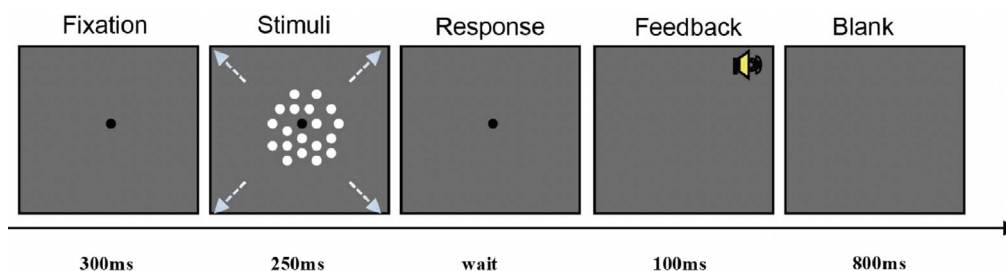

Figure 2. Illustration of a single trial in the 4AFC global motion direction identification experiment. Each trial began with a brief tone, followed by a 250-ms intervals that contained 400 moving dots. Observers indicated the direction of coherent motion (45°, 135°, 225°, and 315°) by pressing the button on keyboard. The proportion of coherent motion could be updated according to subjects' response. Auditory feedback was given after each correct response. A 1-s blank screen was shown between trials.

Starting levels	Observer 1 (%)			Observer 2 (%)			Observer 3 (%)		
	40	80	160	40	80	160	40	80	160
+25%	60.7	100	100	88.0	100	100	93.7	100	100
0%	83.1	100	100	93.0	100	100	95.2	100	100
−25%	84.8	100	100	89.9	100	100	91.9	100	100

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

observers. Specifically, we simulated the even trials with the RSS method. Following Kattner et al. (2017), we simulated 1,000 runs of the experiment (only the even trials) with stimulus intensities in each trial randomly selected from 1,000 possible values ranging from 0.001 to 1. Then, we fit the trial-by-trial data from each run of the RSS procedure with a maximum likelihood method to obtain the best fitting exponential learning curve. See details in Supplementary Appendix B.

In order to measure the accuracy and precision of estimation from the qCD, staircase, and RSS methods, bias, root-mean-square error (RMSE) and standard deviation (*SD*) were computed (see details in Supplementary Appendix C: Evaluation Methods). We tested the qCD and staircase methods with the same starting levels: +25%, 0%, and −25% from the true threshold (i.e., proportion of coherent dots) in the first trial. Because the pattern of results with the three starting levels exhibited a similar trend, we only present the results with the 0% starting level in the main text. The results with the other two starting levels are presented in the Supplementary Appendix D. We did not vary the initial stimulus intensity level in the RSS method because it is randomly selected by the method.

Results

Staircase convergence

Estimated thresholds in the simulated staircase procedure were computed by averaging thresholds in the remaining even number of reversals, after deleting the first four or five reversals in each block. We used three different block sizes—40, 80, and 160 trials per block—in the staircase simulations. The respective simulations are labeled with block size. For example, SC80 denotes the simulation in which the block thresholds were computed every 80 trials (e.g., Trials 2, 4, 6 . . . and 160 for Block 1) and the estimated threshold of Block *b* was used to predict the threshold of trial $(b - 0.5) \times 80$ on the learning curve form the staircase method. Table 2 shows the percentage of staircase runs in the first block of the simulated experiment that had more than five reversals. With a block size of 40 trials per block, we can't reliably estimate the threshold in the first block of the simulated experiment because we didn't obtain a sufficient

number of reversals in some cases, especially when the learning is more rapid (e.g., Simulated Observer 1). We dropped SC40 in subsequent analyses.

Estimated learning curves

The average estimated thresholds in the trial-by-trial and post hoc segment-by-segment analyses from the simulations of the qCD method, and the block-by-block threshold estimates from the simulated staircase method with the 0% starting level relative to the true initial thresholds, are shown in Figure 3. Visual inspection suggests that the estimated post hoc segment-by-segment thresholds from the qCD methods are very close to the true thresholds, and closer than the block-by-block thresholds from the staircase method (SC80 and SC160) and the estimated threshold from the RSS method. In addition, the estimates from the qCD method are more precise than those from the staircase and RSS methods. We quantify these observations next. The estimated thresholds from sets of simulations with starting levels ±25% (above or below) the true initial thresholds are shown in Supplementary Figure S1 of Supplementary Appendix D.

Evaluation: Accuracy and precision

The biases of the estimated thresholds from the qCD and staircase methods with the 0% starting level and the RSS method are shown in Figure 4A. In the first block of SC80 and SC160, there are obvious biases (0.026 and 0.039 for Observer1; 0.018 and 0.042 for Observer2; 0.009 and 0.024 for Observer3, all in $\log_{10}$ units). As the block number increased, the biases from the staircase method significantly decreased and finally reached around 0.007 $\log_{10}$ units. The estimated thresholds from the staircase method with 80 trials/block (SC80) were more accurate than those from SC160, but only in the early phase of learning where learned changes are most rapid; the two block sizes are similar later during the saturated phase of learning.

In the qCD method, the biases of the estimated thresholds from the post hoc segment-by-segment analysis were much smaller than that of the trial-by-trial estimates of thresholds in the early phase of learning for the three observers. Because the post hoc segment-by-segment qCD analysis used information from all the trials, we are more interested in the comparison between the post hoc segment-by-segment qCD and the raw block estimates from the staircase method. Across the learning curve, the RMSE of the estimated post hoc segment-by-segment thresholds from the simulated qCD method was 0.005 for Observer 1, 0.003 for Observer 2, and 0.002 for Observer 3, respectively (all in $\log_{10}$ units). In
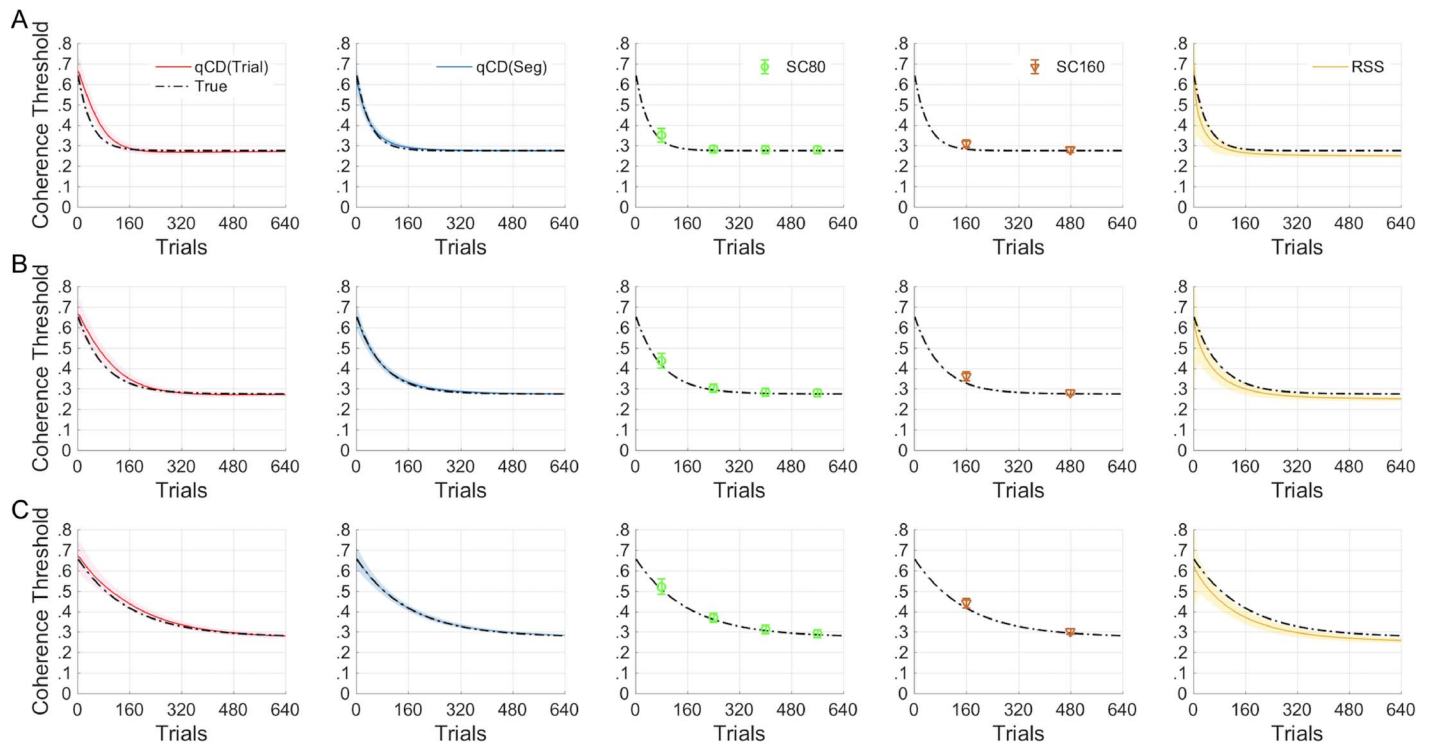
Figure 3. Estimated learning curves in the first 640 trials of simulated Observers 1 (A), 2 (B), and 3 (C) for starting stimulus levels at 0% from the true thresholds in the first trial. Results from the qCD, staircase, and RSS methods are shown. Red and blue lines denote the estimates in trial-by-trial and post hoc segment-by-segment analyses from qCD simulations. Green circles and brown inverted triangles represent the block-by-block thresholds from the staircase method with block sizes of 80 and 160, respectively. Gold lines denote the estimates from the RSS method. The light red and blue shaded areas denote the *SD* in the trial-by-trial and post hoc segment-by-segment qCD analyses, respectively. The light gold shaded areas and error bars denote the *SD* in the RSS, and staircase methods, respectively.

comparison, the RMSE of the estimated block thresholds (e.g., at the points of the empirical thresholds) from SC80 and SC160 was 0.025 and 0.033 for Observer 1, 0.026 and 0.033 for Observer 2, 0.024 and 0.039 for Observer 3 (all in $\log_{10}$ units). The RMSE of the estimated threshold from the RSS method was 0.084, 0.084 and 0.084 $\log_{10}$ units for Observers 1, 2 and 3, respectively.

In summary, the accuracy of the estimated thresholds from the qCD method was much higher than that from both the staircase and RSS methods. The biases of the estimated thresholds for simulations with starting levels ±25% (above or below) the true initial thresholds are shown in Supplementary Figure S2 of Supplementary Appendix D.

The *SD*s of the estimated thresholds from the qCD and staircase methods from simulations with the 0% starting level and the RSS method are shown in Figure 4B. The estimated thresholds in the staircase method with larger block size (e.g., 160 trials) had smaller *SD*s than those with smaller block size because there are more reversals in larger blocks. In the RSS method, the *SD*s started with a large number, then decreased with training trials. In the qCD method, the *SD*s of the

estimated thresholds in the trial-by-trial and post hoc segment-by-segment analyses also significantly decreased as the trial number increased, but the post hoc segment-by-segment analysis provided smaller *SD*s. Furthermore, the *SD*s of the estimated post hoc segment-by-segment thresholds from the qCD method were always considerably smaller than those from SC80, SC160, and the RSS method for the three simulated observers. Averaged across the whole learning curve, the *SD* of the estimated post hoc segment-by-segment thresholds from the qCD method were 0.009, 0.010, and 0.011 for the three simulated observers; the *SD*s of the estimated block thresholds in SC80 and SC160 were 0.029 and 0.020 for Observer 1, 0.029 and 0.020 for Observer 2, and 0.028 and 0.020 for Observer 3; the *SD*s of the estimated thresholds from the RSS method were 0.024, 0.025, and 0.026 for the three simulated observers (all in $\log_{10}$ units, see Table 3). In summary, the precision of the thresholds estimated from the qCD method was much higher than those estimated from the staircase and RSS methods, based on these simulations. The *SD*s of the estimated thresholds with starting levels ±25% (above or below)
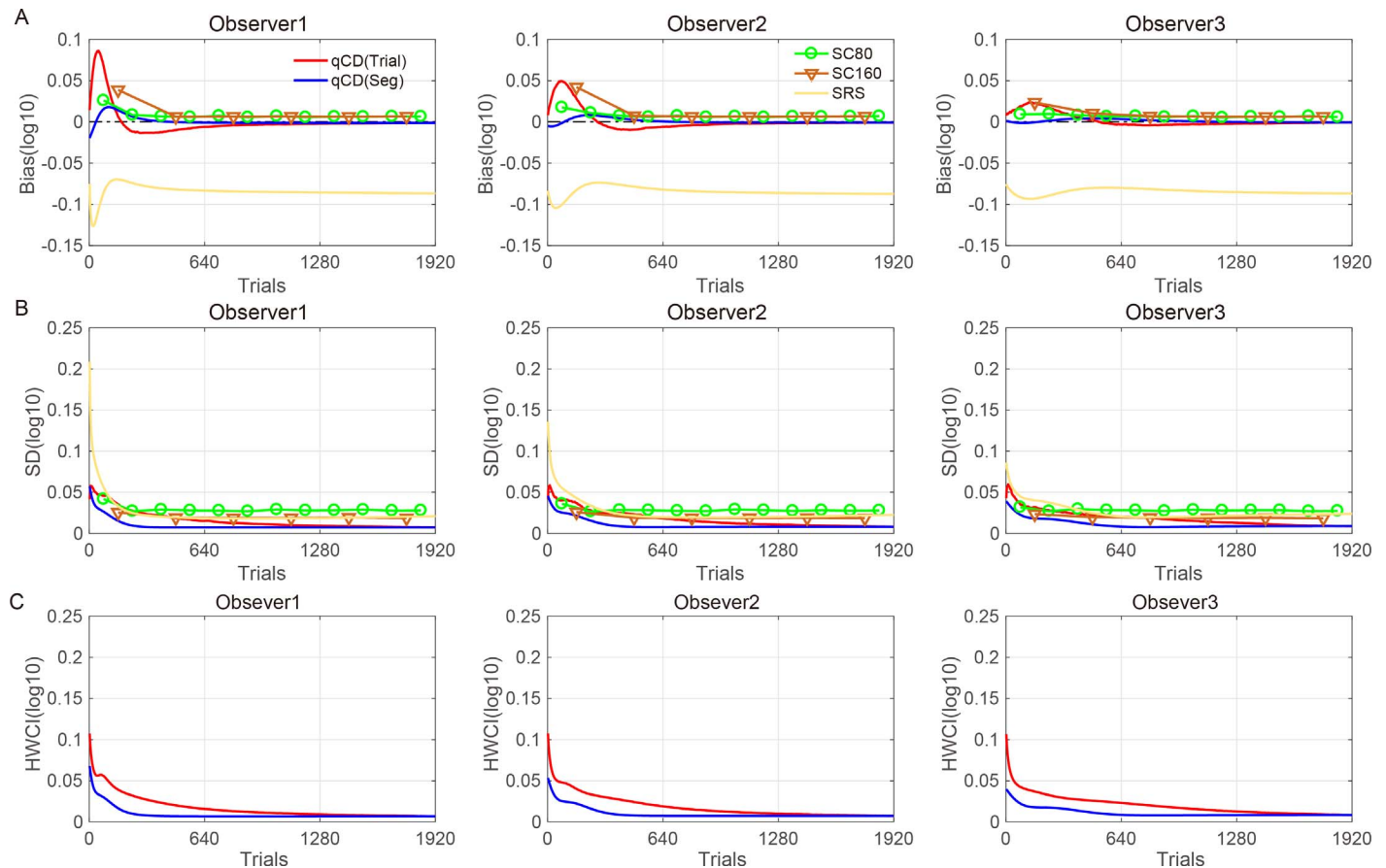
Figure 4. Comparisons of the accuracy and precision of the estimated thresholds from the simulations of the qCD, staircase and RSS methods. The biases (A), *SD*s (B), and 68.2% HWCIs (C) of the three simulated observers for starting levels at 0% from the true threshold in the first trial are shown in separate rows as functions of trial numbers. Red, blue, green, and brown colors denote results from the trial-by-trial qCD analyses, the post hoc segment-by-segment qCD analyses, and the 80- and 160-trial block staircase methods, respectively. The gold color denotes the results from the RSS method.

the true initial thresholds are shown in Supplementary Figure S3 of Supplementary Appendix D.

In Figure 4C, half widths of the 68.2% credible interval (HWCIs) of the thresholds estimated from the qCD method with the 0% starting level decreased across trials. Averaged across the three simulated observers, the 68.2% HWCI of the estimated trial-by-trial thresholds were 0.039, 0.019, and 0.008 $\log_{10}$ units after 160, 640, and 1,920 trials, respectively. Similarly, there is also a monotonic decrease in the 68.2% HWCI from the post hoc segment-by-segment analysis of the estimated learning curves (see the blue line in Figure 4C): The 68.2% HWCI averaged across the three simulated observers was 0.019, 0.008, and 0.008 $\log_{10}$ units after 160, 640, and 1,920 trials, respectively. Detailed results of the 68.2% HWCIs with the three starting levels are summarized in Table 4 and Supplementary Figure S4 of Supplementary Appendix D.

The biases and *SD*s of the estimated parameters of the learning curves from the qCD and staircase methods with the 0% starting level and RSS method are shown in

Figure 5. 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 optimized parameters of the exponential model fit to the threshold estimates. For the RSS method, trial-by-trial data were fit with an exponential function with a maximum likelihood method to obtain the estimated parameters. The biases of the estimated parameters from the qCD method were much smaller than those from the staircase and RSS methods, especially for the simulated Observer 1 with a faster learning parameter. For example, when the time constant was 80 trials (Observer 2), the biases of the estimated λ from the qCD, SC80, SC160, and RSS methods were −0.016, 0.210, 0.490, and −0.096 $\log_{10}$ units, respectively. Similarly, the *SD*s of the parameter estimates from the post hoc segment-by-segment qCD method were much lower than those derived from fitting the exponential to the staircase method thresholds and RSS method. For example, for Observer 2 the *SD*s of the λs estimated from the qCD, SC80, SC160, and RSS methods were 0.081, 0.648,

	Trial 80				Trial 160				Average cross learning curve				
	qCDtrial	qCDseg	SC80	RSS	qCDtrial	qCDseg	SC160	RSS	qCDtrial	qCDseg	SC80	SC160	RSS
Observer 1													
+25%	0.046	0.027	0.042		0.034	0.015	0.025		0.016	0.009	0.030	0.020	
0%	0.046	0.026	0.042	0.033	0.035	0.015	0.026	0.022	0.016	0.009	0.029	0.020	0.024
−25%	0.045	0.025	0.038		0.034	0.014	0.024		0.016	0.009	0.029	0.020	
Observer 2													
+25%	0.039	0.025	0.038		0.035	0.021	0.026		0.017	0.010	0.029	0.020	
0%	0.040	0.025	0.036	0.430	0.037	0.022	0.026	0.027	0.016	0.010	0.029	0.020	0.025
−25%	0.039	0.025	0.035		0.035	0.021	0.025		0.017	0.010	0.029	0.020	
Observer 3													
+25%	0.035	0.021	0.032		0.031	0.018	0.023		0.019	0.011	0.029	0.020	
0%	0.035	0.023	0.032	0.040	0.032	0.018	0.024	0.033	0.018	0.011	0.028	0.020	0.026
−25%	0.035	0.021	0.031		0.031	0.018	0.022		0.018	0.011	0.029	0.020	

Table 3. The standard deviations of the estimated thresholds in the qCD, staircase and RSS methods. *Notes*: Standard deviations estimated at two key points and across the learning curve for the different methods. qCDtrial and qCDseg denote the results from the trial-by-trial and post hoc segment-by-segment qCD simulations, respectively. The 80 and 160 trial points correspond with the trial placement (average trial number) of the first point estimated by the SC80 and SC160 methods. All results are in $\log_{10}$ units. Because only odd number trials in the learning curve were simulated by the qCD method, the Trial 80 and 160 in the qCD method correspond with Trials 79 and 159, which are labeled at 80 and 160 for convenience.

0.884, and 0.259 $\log_{10}$ units, respectively. Note that 0.1, 0.5, and 1 $\log_{10}$ units denote about 25%, 300%, and 1,000% (ratio) deviation from the truth, respectively. Based on the simulations, the qCD method yielded higher accuracy and precision for estimated parameters than the staircase and RSS methods. Furthermore, both the staircase and RSS methods were less effective in estimating the parameters when learning was rapid (Observer 1), while the accuracy and precision on estimated parameters from the qCD method yielded good estimates in all cases (see Tables 5 and 6 for details). Furthermore, the different starting levels were more likely to affect the accuracy of the estimated parameters in the staircase method, while the biases of parameters estimated from the qCD method did not vary much with the starting level. For example, the

biases of the estimated λ from the qCD method with +25%, 0% and −25% starting levels were −0.057, −0.051, and −0.054 $\log_{10}$ units, respectively, but were 0.221, 0.321, and 0.336 $\log_{10}$ units with the SC80 method (see Tables 5 and 6, and Supplementary Figures S5 and S6 of Supplementary Appendix D for details).

The average 68.2% HWCIs of the parameters estimated from the qCD method are shown in Figure 6. The 68.2% HWCIs decreased as the number of trials increased. Averaged across the three simulated observers, the 68.2% HWCI of the estimated λ was 0.114, 0.097, and 0.096 $\log_{10}$ units after 1, 640, and 1,920 trials, respectively. The 68.2% HWCI of the estimated γ was 0.237, 0.156, and 0.115 $\log_{10}$ units after 1, 640, and 1,920 trials, respectively. The 68.2% HWCI of the estimated α was 0.088, 0.027, and 0.009 $\log_{10}$ units after

	Trial 1	Trial 160	Trial 640	Trial 1,280	Trial 1920	Average
Observer 1						
+25%	0.110; 0.070	0.039; 0.017	0.016; 0.007	0.009; 0.007	0.007; 0.007	0.017; 0.009
0%	0.107; 0.068	0.039; 0.017	0.016; 0.007	0.009; 0.007	0.007; 0.007	0.017; 0.009
−25%	0.114; 0.069	0.039; 0.016	0.016; 0.007	0.009; 0.007	0.007; 0.007	0.017; 0.009
Observer 2						
+25%	0.110; 0.053	0.040; 0.022	0.019; 0.007	0.010; 0.007	0.007; 0.007	0.019; 0.010
0%	0.108; 0.053	0.040; 0.022	0.019; 0.007	0.010; 0.007	0.007; 0.007	0.019; 0.010
−25%	0.112; 0.054	0.040; 0.022	0.019; 0.007	0.010; 0.007	0.007; 0.007	0.019; 0.010
Observer 3						
+25%	0.109; 0.040	0.037; 0.018	0.023; 0.009	0.013; 0.008	0.008; 0.008	0.020; 0.011
0%	0.107; 0.040	0.037; 0.018	0.023; 0.009	0.013; 0.008	0.008; 0.008	0.020; 0.011
−25%	0.111; 0.040	0.037; 0.018	0.023; 0.009	0.013; 0.008	0.008; 0.008	0.020; 0.011

Table 4. The half widths of the 68.2% credible interval (HWCIs) of the estimated thresholds from the qCD method. *Notes*: The 68.2% HWCI of the simulated thresholds estimated from the qCD method in the trial-by-trial (the first value in each cell) and post hoc segment-by-segment (the second value in each cell) analyses. All values are in $\log_{10}$ units.
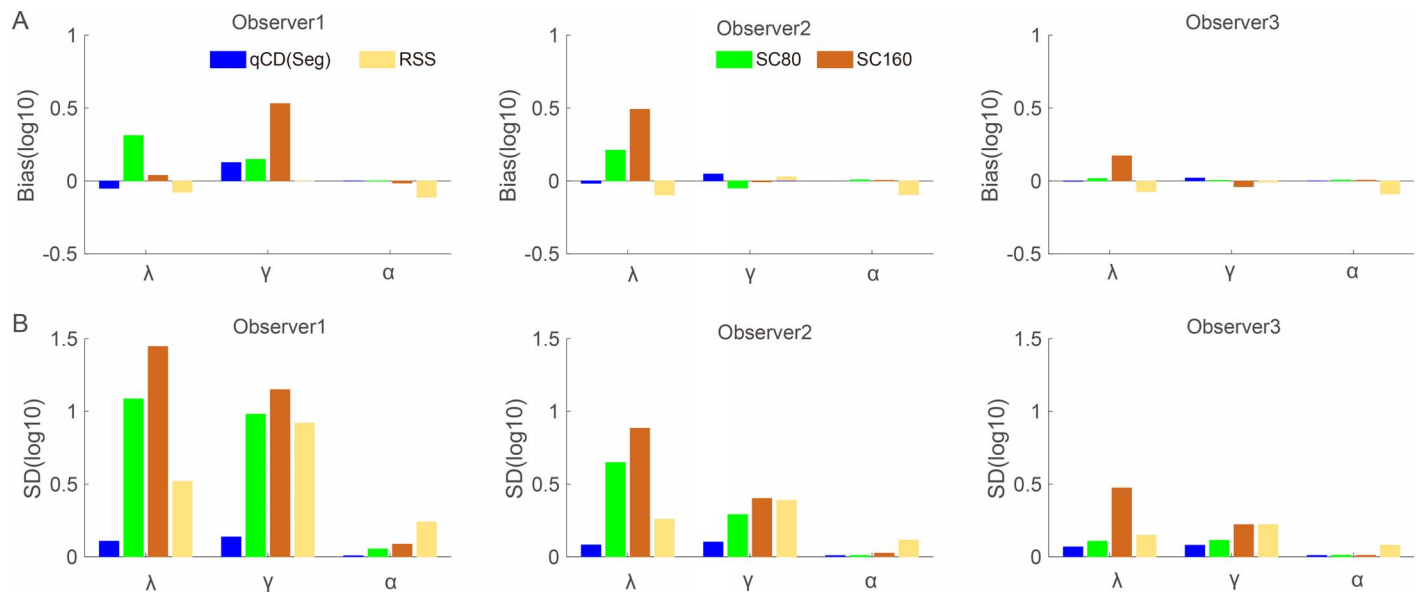
Figure 5. The biases (A) and *SD*s (B) of the estimated parameters (λ, γ, and α) of the three simulated observers from the post hoc segment-by-segment qCD (blue), the SC80 (green), and SC160 (brown) staircase measures, and the RSS method (gold).

1, 640, and 1,920 trials, respectively (see Table 7 and Supplementary Figure S7 of Supplementary Appendix D for details). The results indicate that the qCD method can estimate the parameters of the learning curve with relatively high precision.

The initial threshold (IT = $\lambda + \alpha$) and percent of improvements (PI = $(\lambda + \alpha)/\alpha \times 100\%$) are critical for understanding characteristics of perceptual learning, such as specificity, transfer, and retention. In Figure 7A, the histogram of the estimated IT from the qCD method is distributed symmetrically and tightly around the true IT (=0.66). However, the distributions of the estimated IT from SC80, SC160, and RSS have long

tails in one direction, indicating systematic and sometime large biases, with the corresponding effect on the *SD*. For example, when the time constant was 80 trials (simulated Observer 2), the estimated IT was 0.652 ± 0.069 ($M \pm SD$) from the qCD method, 1.871 ± 2.792 from SC80, 3.818 ± 3.691 from SC160, and 0.674 ± 0.749 from RSS. The distributions of the estimated PI (Figure 7B) from the qCD method were also much narrower and closer to the true PI (=240%) compared to those from SC80, SC160, and RSS. For example, when the time constant was 80 trials (Observer 2), the estimated PI was $237\% \pm 25\%$ ($M \pm SD$) from the qCD method, $665\% \pm 986\%$ from SC80,

	Parameters											
	λ				γ				α			
	qCDseg	SC80	SC160	RSS	qCDseg	SC80	SC160	RSS	qCDseg	SC80	SC160	RSS
Observer 1												
+25%	−0.057	0.221	−0.062		0.131	0.197	0.579		−0.001	−0.107	−0.032	
0%	−0.051	0.312	0.038	−0.078	0.126	0.149	0.531	−0.001	−0.001	−0.002	−0.015	−0.113
−25%	−0.054	0.336	0.025		0.122	0.131	0.526		−0.001	−0.002	−0.018	
Observer 2												
+25%	−0.022	0.017	0.469		0.057	−0.033	0.001		−0.001	0.006	0.001	
0%	−0.016	0.210	0.491	−0.096	0.047	−0.048	−0.007	0.028	−0.001	0.006	0.001	−0.094
−25%	−0.024	0.206	0.446		0.057	−0.048	0.016		−0.001	0.006	0.001	
Observer 3												
+25%	−0.009	0.018	0.172		0.025	0.003	−0.042		−0.001	0.006	0.006	
0%	−0.003	0.016	0.172	−0.074	0.020	0.003	−0.040	−0.009	−0.001	0.006	0.005	−0.090
−25%	−0.008	0.002	0.163		0.025	0.009	−0.039		−0.001	0.006	0.006	

Table 5. Biases of the estimated parameters from the qCD, staircase, and RSS methods. *Notes*: Average biases of the estimated parameters describing the three simulated observers with starting levels at +25%, 0%, and −25% from the true threshold in the first trial, for the qCD (post hoc segment-by-segment), staircase, and RSS methods. All values are in $\log_{10}$ units.

	Parameters											
	λ				γ				α			
	qCDseg	SC80	SC160	RSS	qCDseg	SC80	SC160	RSS	qCDseg	SC80	SC160	RSS
Observer 1												
+25%	0.113	1.237	1.591		0.141	1.053	1.182		0.007	0.307	0.447	
0%	0.108	1.086	1.447	0.519	0.137	0.981	1.150	0.920	0.008	0.054	0.086	0.240
−25%	0.110	1.084	1.595		0.136	0.921	1.141		0.007	0.058	0.087	
Observer 2												
+25%	0.086	0.633	0.911		0.104	0.290	0.078		0.008	0.010	0.078	
0%	0.081	0.648	0.884	0.259	0.102	0.290	0.025	0.389	0.008	0.009	0.025	0.115
−25%	0.086	0.682	0.912		0.105	0.310	0.079		0.008	0.010	0.079	
Observer 3												
+25%	0.063	0.105	0.459		0.079	0.114	0.011		0.009	0.011	0.011	
0%	0.068	0.108	0.473	0.149	0.079	0.113	0.010	0.222	0.010	0.011	0.010	0.080
−25%	0.062	0.108	0.492		0.075	0.120	0.011		0.009	0.011	0.011	

Table 6. Standard deviations (SDs) of the estimated parameters from the qCD, staircase, and RSS methods. *Notes*: Average SDs of the estimated parameters describing the three simulated observers with starting levels at +25%, 0%, and −25% from the true threshold in the first trial for the qCD (post hoc segment-by-segment), staircase, and RSS methods. All values are in $\log_{10}$ units.

1,367% ± 1,327% from SC160, and 335% ± 1424% from RSS. The estimated PI from the staircase and RSS methods had large SDs and deviated from the true 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 and RSS method. The means and SDs of the estimated IT and PI of the three observers with 3 starting levels are summarized in Table 8 (see Supplementary Figure S8 and S9 of Supplementary Appendix D for the distributions with +25% and −25% starting levels).

Psychophysical validation

A psychophysical experiment was conducted to evaluate and compare the performance of the qCD and staircase methods in measuring the learning curve in a global motion direction identification task with a

relatively slow learning rate. As described in the Introduction, we specifically chose the global motion direction task because it usually exhibits slow learning. The simulation study found that the staircase method was less able to track the detailed time course of fast perceptual learning, so the slow learning rate made it possible for the staircase method to obtain a relatively accurate learning curve and facilitated the comparison between the two methods.

Method

Observers

Five observers (26.40 ± 1.48 years) with normal or corrected-to-normal vision participated in this study. They were all naive to perceptual learning and psychophysical studies. The study protocol was approved by the Institutional Review Board of The Ohio State University. Written informed consent was ob-
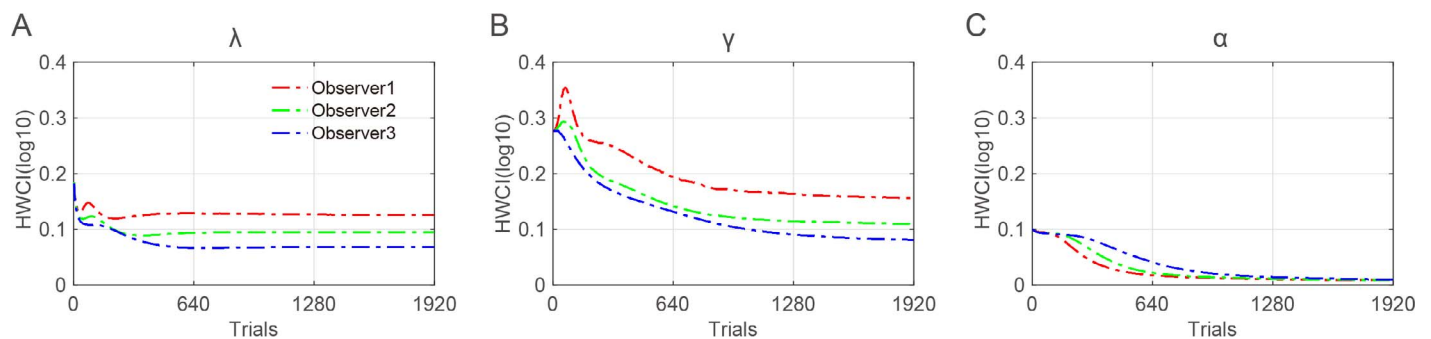

Figure 6. The average 68.2% HWCIs of the λ (A), γ (B), and α (C) parameters estimated from the trial-by-trial qCD analysis of the three simulated observers with starting levels at 0% from the true threshold in the first trial. The red, green, and blue dashed lines denote Observers 1, 2, and 3, respectively.

	Parameters								
	λ			γ			α		
	Trial 1	Trial 640	Trial 1,920	Trial 1	Trial 640	Trial 1,920	Trial 1	Trial 640	Trial 1,920
Observer 1									
+25%	0.188	0.133	0.130	0.277	0.207	0.163	0.101	0.018	0.009
0%	0.184	0.129	0.126	0.276	0.195	0.156	0.101	0.018	0.009
−25%	0.189	0.131	0.129	0.276	0.196	0.160	0.101	0.018	0.009
Observer 2									
+25%	0.189	0.094	0.094	0.277	0.145	0.110	0.101	0.023	0.009
0%	0.184	0.094	0.095	0.276	0.142	0.110	0.101	0.022	0.009
−25%	0.190	0.095	0.096	0.275	0.147	0.111	0.101	0.023	0.009
Observer 3									
+25%	0.188	0.067	0.069	0.277	0.134	0.082	0.101	0.041	0.010
0%	0.182	0.067	0.068	0.276	0.132	0.081	0.100	0.041	0.010
−25%	0.189	0.068	0.069	0.276	0.134	0.081	0.101	0.041	0.010

Table 7. The 68.2% HWCIs of the estimated parameters from the qCD method. *Notes*: Average 68.2% HWCIs of the estimated thresholds of the three observers with starting levels at +25%, 0%, and −25% from the true threshold in the first trial from trial-by-trial qCD analysis. All values are in $\log_{10}$ units.

tained from each observer before the study. The work was carried out in accordance with the Declaration of Helsinki.

Apparatus

The study was conducted on a PC computer running MATLAB (MathWorks, Natick, MA) programs with PsychToolbox extensions (Brainard, 1997; Pelli, 1997). Stimuli were displayed on a Dell CRT monitor with a 1280×1024 pixel resolution, 85 Hz frame rate, and a background luminance of 40 cd/m^2. Observers placed their head on a chin rest and viewed the displays binocularly. The display subtended $27.8° \times 21.6°$ at a viewing distance of 0.69 m.

Design

Each of the five observers performed 960 odd numbered trials (e.g., 1, 3, 5, . . . and 1,919) with the qCD method interleaved with 960 even numbered trials
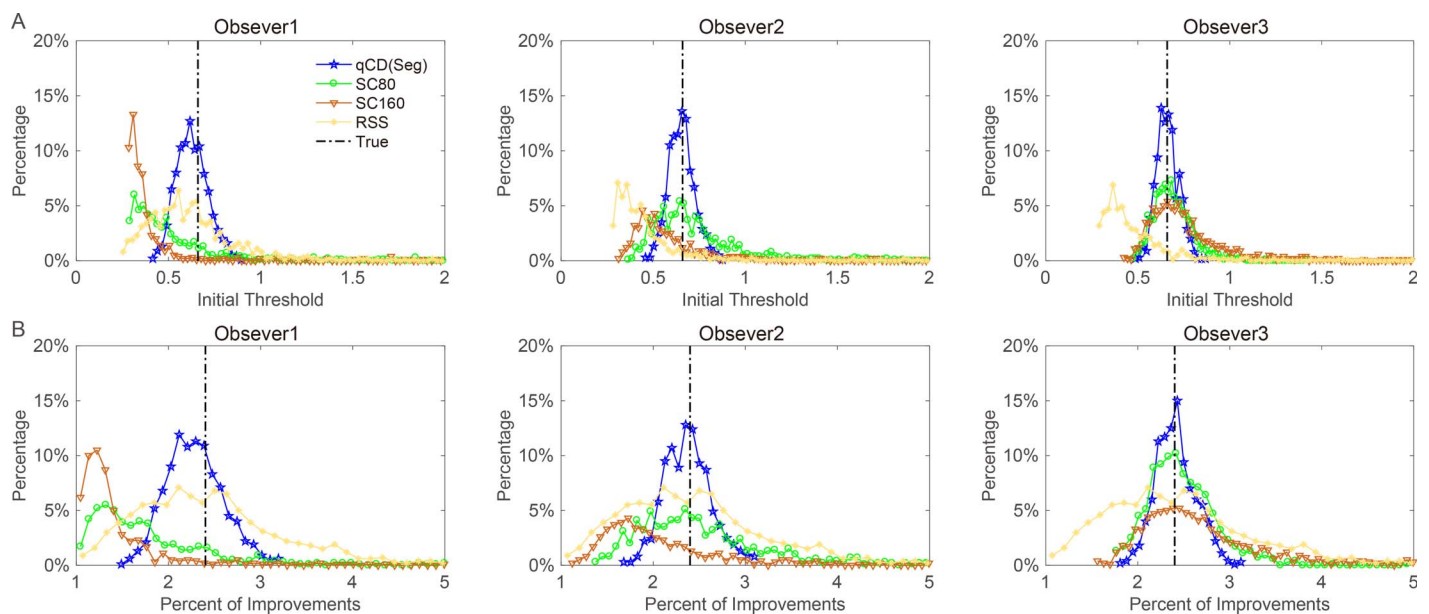

Figure 7. The distributions of the initial threshold (A) and percent of improvements (B) from the qCD, staircase, and RSS methods with starting levels at 0% from the true threshold in the first trial. Blue, green, brown, and gold colors denote the results from the qCD (post hoc segment-by-segment), staircase method with block sizes of 80 and 160, and RSS method, respectively. Black dashed lines denote the true value.

	Initial Threshold				Percent of Improvements			
	qCDseg	SC80	SC160	RSS	qCDseg	SC80	SC160	RSS
Observer 1								
+25%	0.623 (0.086)	3.359 (3.672)	2.600 (32.857)	0.963 (1.719)	227 (31)	>9999	>9999	667 (2919)
0%	0.627 (0.084)	3.549 (3.756)	2.777 (3.004)		229 (31)	1272 (1339)	1007 (1075)	
−25%	0.625 (0.083)	3.695 (3.789)	2.794 (2.968)		228 (30)	1326 (1353)	1011 (1058)	
Observer 2								
+25%	0.648 (0.073)	1.715 (2.655)	3.790 (3.672)	0.674 (0.749)	236 (26)	612 (945)	1395 (1705)	335 (1424)
0%	0.652 (0.069)	1.871 (2.792)	3.818 (3.691)		237 (25)	665 (986)	1367 (1320)	
−25%	0.646 (0.072)	1.949 (2.911)	3.696 (3.719)		236 (26)	695 (1035)	1354 (1628)	
Observer 3								
+25%	0.655 (0.055)	0.692 (0.105)	1.345 (1.975)	0.637 (0.456)	239 (20)	248 (363)	479 (698)	292 (1112)
0%	0.662 (0.060)	0.691 (0.107)	1.372 (2.044)		241 (22)	248 (365)	489 (721)	
−25%	0.656 (0.055)	0.678 (0.107)	1.396 (2.128)		239 (20)	243 (369)	496 (749)	

Table 8. Mean and standard deviations of the estimated initial threshold (IT) and percent of improvements (PI) from the qCD and staircase methods. Notes: Mean and standard deviation of the estimated IT and PI from the qCD (post hoc segment-by-segment), staircase and RSS methods.

(e.g., 2, 4, 6, . . . and 1,920) from a three-down/one-up staircase. The total of 1,920 trials of training were evenly divided into three sessions on different days, each of which consisted of four blocks of 160 trials. (Note: observer N3 only finished the first 11 blocks [880 trials from the qCD method interleaved with 880 trials from the staircase method]). The block estimates from the staircase method were computed from every 80 trials (SC80; e.g., Trials 2, 4, 6. . . and 160 for Block 1).

Procedure

In each trial, a random dot kinematogram (RDK) of 400 moving dots (dot size: 0.18° × 0.18°; speed: 10°/s) in an 8° diameter circular aperture (Figure 2) was presented for 250 ms after a brief tone. The signal dots in the RDK could move in one of four directions (45°, 135°, 225°, or 315°), while the noise dots moved in random directions. The proportion of motion coherence was determined by the qCD or the staircase method. A small dark fixation point (0.15°) was always displayed in the center of the display. Observers were asked to judge the global direction of the RDK by pressing buttons on a computer keyboard. Auditory feedback was given on correct responses during training.

Results

The estimated trial-by-trial and post hoc segment-by-segment thresholds from the qCD method and the estimated block-by-block thresholds from the staircase method are shown in Figure 8. The estimated dynamic range (λ) from the qCD method was computed from

both trial-by-trial and post hoc segment-by-segment analyses. In the trial-by-trial analysis of the qCD data, the estimated thresholds across all the trials were first fit by an exponential function, the estimated dynamic range was obtained by the parameters of the best fitting model. Averaged across five observers, the dynamic range was 0.272 ± 0.033 ($M \pm SE$). Similarly, the averaged dynamic range across five observers in the post hoc segment-by-segment analysis was 0.311 ± 0.051 ($M \pm SE$). There was no significant difference between the estimated dynamic ranges from the trial-by-trial and post hoc segment-by-segment analyses of the qCD data, $t(4) = -1.823$, $p = 0.142$.

For the staircase method, block-by-block thresholds from the SC80 procedure were obtained. Dynamic range was calculated in two ways: (a) Using the thresholds in the first and last blocks as the initial and final thresholds, the average dynamic range across the five observers was 0.308 ± 0.057 ($M \pm SE$); (b) Based on the estimated thresholds of the first and last trial calculated from the best fitting exponential model to the block-by-block thresholds, the averaged dynamic range was 0.393 ± 0.072 ($M \pm SE$). The estimated PIs in this two different ways were significant different, $t(4) = -3.28$, $p = 0.031$.

Precision of the estimated thresholds from the qCD method

Figure 9 shows that the 68.2% HWCI of the estimated thresholds from the qCD method decreased with the number of training trials in the trial-by-trial analysis: Averaged across five observers, the 68.2% HWCI of the estimated threshold was 0.030 ± 0.001, 0.022 ± 0.001, and 0.011 ± 0.000 ($M \pm SE$) $\log_{10}$ units after 160, 640, and 1,760 trials. Similarly, the 68.2%
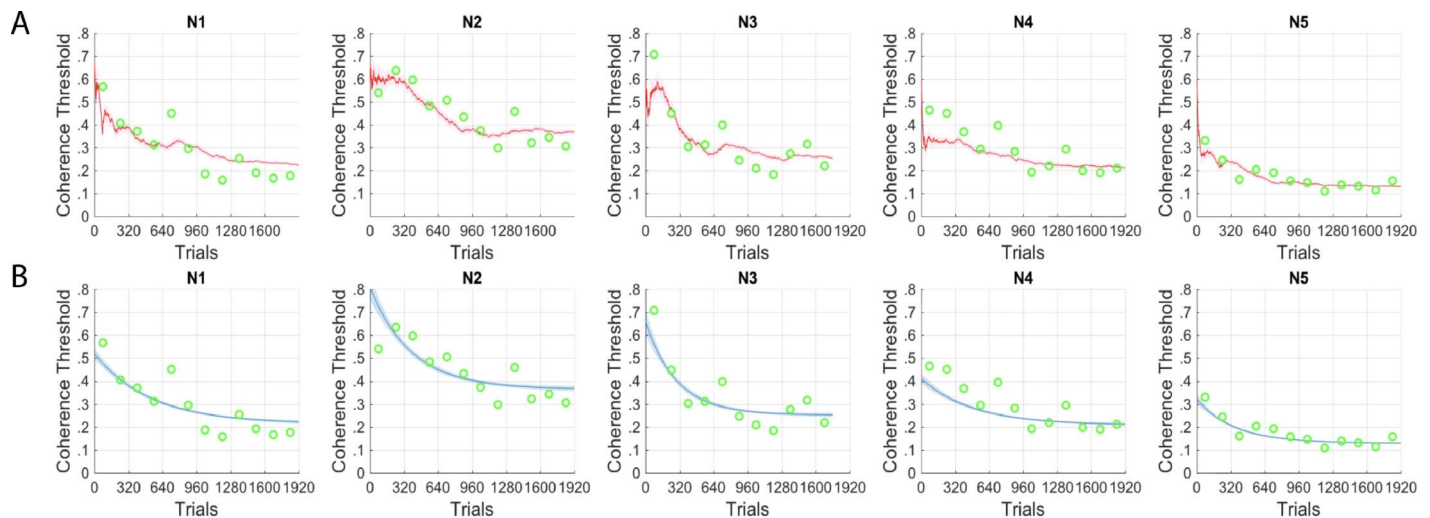
Figure 8. Learning curves for individual observers from the psychophysical experiment. (A) Trial-by-trial estimated thresholds (red line) with 68.2% HWCI (light red shade) from the trial-by-trial qCD method and block-by-block thresholds (green circles) from the staircase method (SC80). (B) Estimated thresholds from the post hoc segment-by-segment qCD analysis (blue line) with 68.2% HWCI (light blue shade).

HWCI of the estimated thresholds from the post hoc segment-by-segment analysis also decreased monotonically with trial number: The average 68.2% HWCI was 0.017 ± 0.001, 0.012 ± 0.001, and 0.013 ± 0.001 $\log_{10}$ units ($M \pm SE$) after 160, 640, and 1,760 trials, respectively. These results indicated that the qCD method could precisely estimate the learning curve in the global motion direction identification task.

The 68.2% HWCI of the estimated parameters of the exponential learning curve for the five observers are shown in Figure 10. Obviously, the precision of the estimated parameters in the trial-by-trial analysis increased with the number of training trials. Across the five observers, the average 68.2% HWCI of λ started at 0.197 in the first trial, and decreased to 0.130 after 320 trials, to 0.083 after 640 trials, and 0.052 after 1280 trials; the average 68.2% HWCI of γ started at 0.277 in the first trial, and decreased to 0.171 after 320 trials, to 0.121 after 640 trials, and 0.087 after 1,280 trials; the average 68.2% HWCI of α started at 0.101 in first trial, and decreased to 0.074 after 320 trials, to 0.063 after

640 trials, and 0.033 after 1,280 trials, all in $\log_{10}$ units. These results indicated that the qCD method could precisely estimate parameters of perceptual learning with low speed.

Correlation between the qCD and staircase method

To compute the correlation of the thresholds estimated by the qCD and staircase methods, we carried out a special procedure across observers to eliminate the dependency of estimated thresholds across learning trials (Hou, Lesmes, Bex, Dorr, & Lu, 2015): (1) Estimated thresholds after 40, 120, 200, 280, 360, 440, 520, 600, 680, 760, 840, and 920 ($block^{th} \times 80 - 40$) qCD trials were taken from the post hoc segment-by-segment analysis. Note that since the values from this analysis are based on fitting an exponential function over all values, the thresholds are not independent for values close in time. (2) The block-by-block thresholds from the 80-trial staircase method were estimated by averaging the reversals in each block.
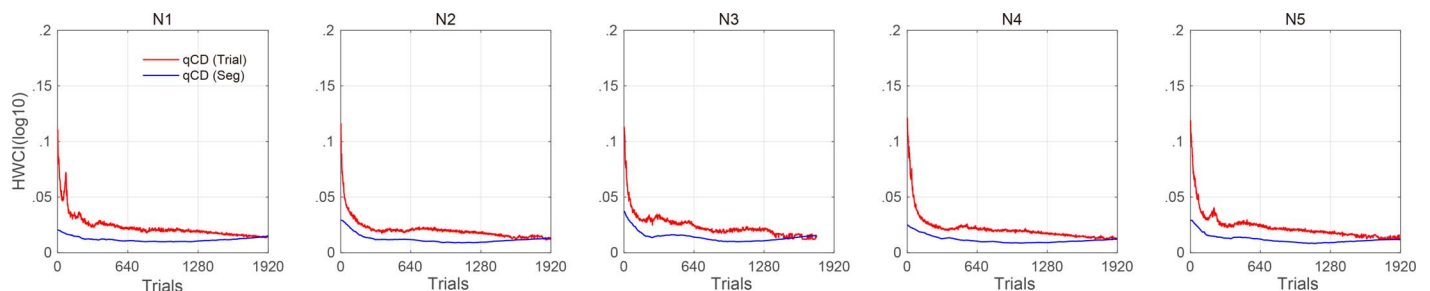

Figure 9. The 68.2% HWCI of the estimated thresholds from the trial-by-trial (red) and post hoc segment-by-segment (blue) qCD analysis as functions of trial numbers for individual subjects.
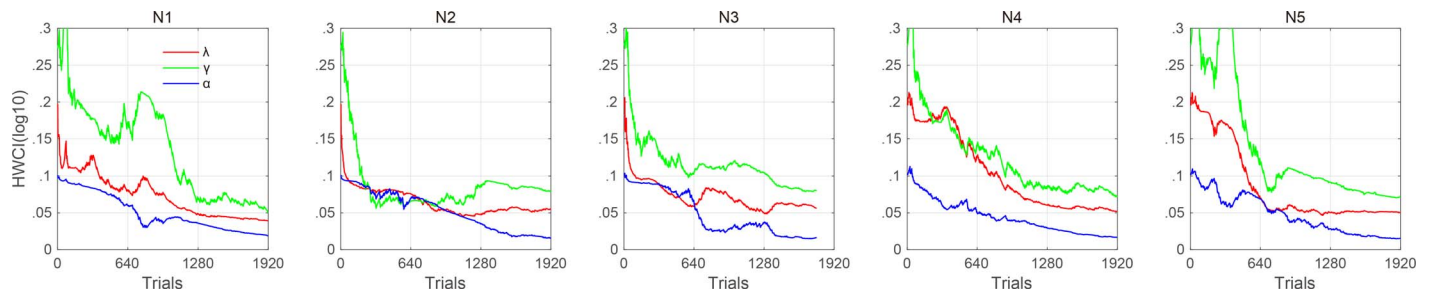
Figure 10. The 68.2% HWCI of the estimated parameters from the trial-by-trial qCD analysis as functions of trial numbers for individual subjects. Red, green, and blue colors represent λ, γ, and α, respectively.

(3) We randomly sampled one pair of the estimated thresholds from the two methods in the same block for Observer N1, and repeated this process for the other four observers without sampling the same block between any two observers. (4) The Pearson correlation coefficient was calculated between the five pairs of estimated thresholds from the two methods for the five observers. (5) This procedure was repeated 1,000 times, and the average Pearson correlation coefficient was obtained. Because the five pairs of estimated thresholds in Step 3 were from different observers, this procedure ensures that the pairs that enter into the Pearson correlation calculation (Step 4) were independent. Figure 11 shows results from 1,000 runs of the procedure. The average Pearson correlation coefficient was 0.879 ± 0.126 ($M \pm SD$). We also computed the RMSE between the estimated thresholds from the two methods. The RMSE was 0.086 ± 0.024 $\log_{10}$ units. Results from both the correlation analysis and RMSE
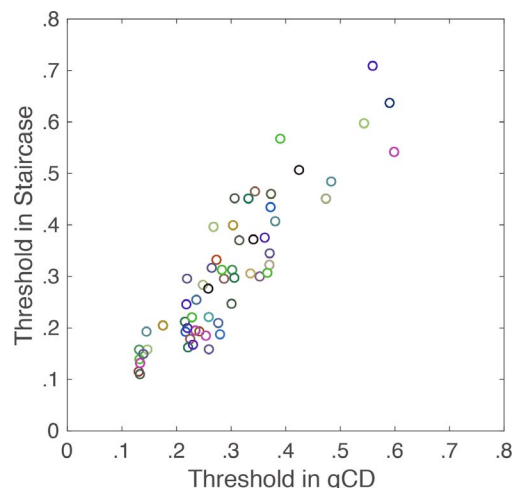

Figure 11. The relationship between thresholds estimated from the post hoc segment-by-segment qCD method and the staircase (post hoc SC80) method from decorrelated subsamples of the data; see the text for explanations and reports of the average correlation. Different colors denote sampled thresholds from different runs of the procedure.

calculation suggest that the estimated thresholds from the two methods matched quite well.

Discussion

In this study, we systematically examined the qCD, staircase, and RSS methods in assessing the learning curve in a 4AFC global motion direction identification task in both simulations and a corresponding psychophysical validation experiment. In both cases, trials from the qCD method were alternated with trials from a staircase or RSS procedure (odd and even trials, respectively). The simulations showed that the estimated learning curves from both the trial-by-trial and post hoc segment-by-segment analyses of the qCD method were more accurate and precise than those obtained from staircase methods using either 80 or 160 trial blocks and the RSS method. The differences were largest for the dynamic range (λ) and time constant (γ) of learning.

In the psychophysical experiment, we chose a global motion direction identification task with a slow learning rate so that the staircase method (80 trial block) could yield reasonably accurate estimates of the learning curve, which could then be compared to the estimates from the qCD methods for the purposes of validation. We found that 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.

During perceptual learning, the perceptual threshold may in many cases change continuously (Lu et al., 2011; Mazur & Hastie, 1978; Petrov et al., 2005), potentially even within each measurement block and especially in the early phase of learning (Badiru, 1992; Dosher & Lu, 2007; Heathcote, Brown, & Mewhort, 2000). Our simulations showed that the staircase method, providing block-by-block threshold estimates, did not track the detailed time course of perceptual learning as well as the qCD method, especially when

the learning rate is fast. Furthermore, the typical data analytical approach that fits a parametric function to the estimated learning curve from the staircase method sometimes yielded biased and imprecise estimates of the initial threshold and the percent improvement in perceptual learning. Similarly, the estimated learning curves and learning parameters from the RSS method were also quite inaccurate and imprecise.

In the qCD method, the learning curve is parameterized as an exponential function with joint probability distributions of all three parameters. Optimizing the expected information gain in the next trial in the adaptive Bayesian framework (Kontsevich & Tyler, 1999; Lesmes et al., 2006; Lesmes et al., 2010; Watson & Pelli, 1983), the method selects the optimal stimulus in each trial—the stimulus that together with the observer's response will provide more new information about the parameter distribution. Using the posterior distribution of the parameters following each trial, the qCD method provides a detailed trial-by-trial estimate of perceptual sensitivity and therefore a detailed time course of perceptual sensitivity change. Both the simulations and psychophysical experiment showed that the estimated learning curves and their parameters from the qCD method were quite accurate and precise.

Like any adaptive procedure such as the often-used staircase method, the stimulus sequences for individual observers in the qCD procedure are different from one another. What these methods control is the range of performance levels (i.e., near threshold) throughout the experiment and across observers. When we consider keeping the learning experience the same across observers, we can either match stimulus sequences (but different performance levels) or performance levels (but different stimulus sequences), but not both. In the current study, we matched the performance levels but used different stimulus sequences to keep the learning experience the same in terms of performance level across observers.

In the current implementation of the qCD method, the same prior is used across observers. To investigate the impact of the initial stimulus intensity, we have in fact artificially "fixed" the stimulus intensity in the first trial. It is true that with the same prior, all participants will be tested with the same stimulus intensity in trial 1. In the trial-by-trial analysis, the estimated threshold is determined by both the stimulus intensity and the participant's response. The estimated threshold would be the same if all the participants make the same response. In the segment-by-segment analysis, the estimated threshold on trial 1 is determined by all the stimuli and responses in the whole segment. Unless the responses of two participants were identical throughout the entire segment, the probability of obtaining the same estimated threshold on trial 1 is very small.

One potential future application of the qCD method in perceptual learning is to improve the estimate of the transfer/specificity index. Although specificity has been taken as the trademark of perceptual learning (Ball & Sekuler, 1982, 1987; Fahle, 2004; Fahle & Morgan, 1996; Fiorentini & Berardi, 1980; Karni & Sagi, 1991; Maehara & Goryo, 2003; Saffell & Matthews, 2003), many studies have found that the phenomenon is graded (Ahissar & Hochstein, 1997; J. Huang, Liang, Zhou, & Liu, 2017; Jeter et al., 2009; Z. Liu & Weinshall, 2000; Xiao et al., 2008). To quantify the degree of specificity or transfer, a transfer index is often computed based on the initial and final thresholds in the learning phase and the initial threshold in the transfer phase (Ahissar & Hochstein, 1997; Dosher & Lu, 2007; C.-B. Huang et al., 2012; Jeter, Dosher, Liu, & Lu, 2010). On the other hand, except for boosting the immediate performance of a new task, generalization may reflect on a faster learning rate of a new task (Kattner, Cochrane, Cox, Gorman, & Green, 2017; Z. Liu & Weinshall, 2000). Our simulations showed that the estimated initial threshold, percent of improvements, and time constant from the staircase method could produce substantial biases and larger *SD*s, and if used, might yield more imprecise estimates of the transfer index. On the other hand, the qCD method might yield accurate and precise estimates of the initial threshold, percent of improvements, and time constant, and therefore much improved estimates of the transfer index.

In perceptual learning studies, how to optimize learning and generalization of learning are very important questions (Dosher & Lu, 2017). Many studies in perceptual learning have shown that including easy trials could improve learning (Lin, Dosher, & Lu, 2017; Liu, Lu, & Dosher, 2012; Rubin, Nakayama, & Shapley, 1997). In developing the qCD method, our focus has primarily been on measuring the time course of perceptual sensitivity change. One possible application of the qCD method in perceptual learning is to separate measurement of the learning curve (which often involve stimulus intensities near threshold) from training itself (which could use high intensity stimuli for example) because, with the more efficient measurement, we don't need to devote all the trials in a study into measuring the learning curve. In this respect, the RSS method, which produced a certain percentage of the easy stimuli, may be more likely to improve perceptual learning compared with setting all stimulus at the threshold. The advantages and disadvantages of these methods may depend on researchers' specific needs and questions.

An exponential functional form was used to model the learning curve in the qCD method based on some empirical evidence from the literature concerning the form of learning curves for individual observers

(Dosher & Lu, 2007). Others have suggested different forms of learning curves, such as learning curves with fast and slow phases (Fahle et al., 1995; Poggio et al., 1992), cascade exponential functions (Ahissar & Hochstein, 1997), or power functions (Anderson & Tweney, 1997; Heathcote et al., 2000). (Note that average learning curves generally follow the power function even when the learning curves of individual observes are exponential.)

We can address the functional form issue in several different ways. First, in the trial-by-trial procedure of the qCD method, the posterior distribution of the exponential function changes from trial to trial. This has allowed us, in another application, to estimate perceptual sensitivity changes consisting of a cascade of exponential functions (Zhao et al., 2019). Second, we can also replace the exponential function in the qCD method with other functional forms that are required for a particular learning task. This requires a careful examination of the empirical literature and/or collection of pilot data. Third, we have also been developing a nonparametric version of the qCD method that does not depend on the particular functional form.

As stated in Zhao et al. (2019), the applicability of the qCD method depends on its rate of posterior convergence relative to the rate of the to-be-estimated perceptual sensitivity change. In order to obtain accurate and precise estimates of the time course of perceptual sensitivity change, the qCD should only be used to measure the time course of perceptual sensitivity changes that are slower or at most compatible to the convergence rate of the joint posterior distribution of the parameters of the time course modeled in the qCD procedure. The convergence of the posterior distribution depends on many factors, including prior distribution, stimulus selection, model complexity, and experimental design. Successful implementation and optimization of the qCD would require careful consideration of all these factors in each application on a case-by-case basis, perhaps based on pilot data and simulation. For example, in the qCD, the prior distribution is constructed based on *a priori* knowledge about the time course of perceptual sensitivity change for the studied population. The more informative the prior is—that is, the more we know about the properties of the population under study— the faster the posterior converges (Gu et al., 2016; Kim, Pitt, Lu, Steyvers, & Myung, 2014).

In conclusion, we have implemented the qCD method to assess the time course of perceptual sensitivity change in perceptual learning. In both simulations and a psychophysics experiment, we demonstrated that the method can provide an accurate and precise assessment of the learning curve in a 4AFC global motion direction identification task.

Keywords: perceptual learning, qCD, staircase, precision, accuracy

References

Ahissar, M., & Hochstein, S. (1996). Learning pop-out detection: Specificities to stimulus characteristics. *Vision Research, 36*(21), 3487–3500.

Ahissar, M., & Hochstein, S. (1997, May 22). Task difficulty and the specificity of perceptual learning. *Nature, 387*(6631), 401–406.

Anderson, R. B., & Tweney, R. D. (1997). Artifactual power curves in forgetting. *Memory & Cognition, 25*(5), 724–730.

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, November 12). A specific and enduring improvement in visual motion discrimination. *Science, 218*(4573): 697–698.

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

Ball, K., Sekuler, R., & Machamer, J. (1983). Detection and identification of moving targets. *Vision Research, 23*(3), 229–238.

Brainard, D. (1997). Psychophysics software for use with MATLAB. *Spatial Vision, 10*, 433–436.

Crist, R. E., Li, W., & Gilbert, C. D. (2001). Learning to see: Experience and attention in primary visual cortex. *Nature Neuroscience, 4*(5), 519–525.

Donovan, I., Szpiro, S. F., & Carrasco, M. (2015). Exogenous attention facilitates location transfer of perceptual learning. *Journal of Vision*, *15*(10):11, 1–16, https://doi.org/10.1167/15.10.11. [PubMed] [Article]

Dosher, B. A., & Lu, Z.-L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceedings of the National Academy of Sciences*, *95*(23), 13988–13993.

Dosher, B. A., & Lu, Z.-L. (1999). Mechanisms of perceptual learning. *Vision Research*, *39*(19), 3197–3221.

Dosher, B. A., & Lu, Z.-L. (2005). Perceptual learning in clear displays optimizes perceptual expertise: Learning the limiting process. *Proceedings of the National Academy of Sciences, USA*, *102*(14), 5286–5290.

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. A., & Lu, Z.-L. (2017). Visual perceptual learning and models. *Annual Review of Vision Science*, *3*(1), 343–363.

Fahle, M. (2004). Perceptual learning: A case for early selection. *Journal of Vision*, *4*(10):4, 879–890, https://doi.org/10.1167/4.10.4. [PubMed] [Article]

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*, *6*(3), 292–297.

Fiorentini, A., & Berardi, N. (1980, September 4). Perceptual learning specific for orientation and spatial frequency. *Nature*, *287*(5777): 43–44.

Ghose, G. M., Yang, T., & Maunsell, J. H. (2002). Physiological correlates of perceptual learning in monkey V1 and V2. *Journal of Neurophysiology*, *87*(4), 1867–1888.

Gold, J., Bennett, P., & Sekuler, A. (1999, November 11). Signal but not noise changes with perceptual learning. *Nature*, *402*(6758), 176–178.

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

Green, C. S., Banai, K., Lu, Z.-L., & Bavelier, D. (2018). Perceptual learning. In *Stevens' handbook of experimental psychology and cognitive neuroscience* (pp. 1–47). American Cancer Society. New York, NY: John Wiley & Sons, Inc.

Gu, H. R., Kim, W., Hou, F., Lesmes, L. A., Pitt, M. A., Lu, Z.-L., & Myung, J. I. (2016). A hierarchical Bayesian approach to adaptive vision testing: A case study with the contrast sensitivity function. *Journal of Vision*, *16*(6):15, 1–17, https://doi.org/10.1167/16.6.15. [PubMed] [Article]

Hawkey, D. J. C., Amitay, S., & Moore, D. R. (2004). Early and rapid perceptual learning. *Nature Neuroscience*, *7*, 1055.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, *7*(2), 185–207.

Herzog, M. H., & Fahle, M. (1997). The role of feedback in learning a vernier discrimination task. *Vision Research*, *37*(15), 2133–2141.

Hou, F., Lesmes, L. A., 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., Lu, Z.-L., & Dosher, B. A. (2012). Co-learning analysis of two perceptual learning tasks with identical input stimuli supports the reweighting hypothesis. *Vision Research*, *61*, 25–32.

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*, *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. *The Journal of Neuroscience*, *34*(25), 8423–8431.

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]

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences*, *88*(11), 4966–4970.

Kattner, F., Cochrane, A., Cox, C. R., Gorman, T. E., & Green, C. S. (2017). Perceptual learning generalization from sequential perceptual training as a change in learning rate. *Current Biology*, *27*(6), 840–846.

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]

Kim, W., Pitt, M. A., Lu, Z.-L., Steyvers, M., & Myung, J. I. (2014). A hierarchical adaptive approach to optimal experimental design. *Neural Computation*, *26*(11), 2465–2492.

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

Kumar, T., & Glaser, D. A. (1993). Initial performance, learning and observer variability for hyperacuity tasks. *Vision Research*, *33*(16), 2287–2300.

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. (2015). Developing Bayesian adaptive methods for estimating sensitivity thresholds (d′) in Yes-No and forced-choice tasks. *Frontiers in Psychology*, *6*, 1070.

Levi, D. M., & Polat, U. (1996). Neural plasticity in adults with amblyopia. *Proceedings of the National Academy of Sciences*, *93*(13), 6830–6834.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, *49*(2), 467–477.

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]

Lin, Z., Dosher, B. A., & Lu, Z.-L. (2017). Mixture of easy trials enables transient and sustained perceptual improvements through priming and perceptual learning. *Scientific Reports*, *7*(1): 7421.

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, 1–14, https://doi.org/10.1167/10.10.29. [PubMed] [Article]

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*(61), 15–24.

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

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. A. (2013). *Visual psychophysics: From laboratory to theory*. Cambridge, MA: The 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.

Lu, Z.-L., Lin, Z.-C., & Dosher, B. A. (2016). Translating perceptual learning from the laboratory to applications. *Trends in Cognitive Sciences*, *20*(8), 561–563.

Lu, Z.-L., Liu, J., & Dosher, B. A. (2010). Modeling mechanisms of perceptual learning with augmented Hebbian re-weighting. *Vision Research*, *50*(4), 375–390.

Lu, Z.-L., Zhang, P., Zhao, Y., & Dosher, B. A. (2018). Evaluating the performance of the staircase and quick change detection methods in measuring perceptual learning. *Journal of Vision*, *18*(10): 256, https://doi.org/10.1167/18.10.256. [Abstract]

Maehara, G., & Goryo, K. (2003). Perceptual learning in visual backward pattern masking. *Perceptual and Motor Skills*, *97*(3f), 1137–1149.

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

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, https://doi.org/10.1523/JNEUROSCI.3002-07.2007.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.

Pelli, D. G., & Bex, P. (2013). Measuring contrast sensitivity. *Vision Research*, 90, 10–14.

Pelli, D. G.,& Farell, B. (1995). Psychophysical methods. In: M. Bass, E. W. Van Stryland, D. R. Williams, & W. L. Wolfe (Eds.), *Handbook of Optics, 2nd ed., I* (pp. 29.21-29.13). New York, NY: McGraw-Hill.

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.

Poggio, T., Fahle, M., & Edelman, S. (1992, May 15). Fast perceptual learning in visual hyperacuity. *Science*, 256(5059), 1018–1021.

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, USA*, 101(17), 6692–6697.

Rubin, N., Nakayama, K., & Shapley, R. (1997). Abrupt learning and retinal size specificity in illusory-contour perception. *Current Biology*, 7(7), 461–467.

Saarinen, J., & Levi, D. M. (1995). Perceptual learning in vernier acuity: What is learned? *Vision Research*, 35(4), 519–527.

Saffell, T., & Matthews, N. (2003). Task-specific perceptual learning on speed and direction discrimination. *Vision Research*, 43(12), 1365–1374.

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

Sasaki, Y., Nanez, J. E., & Watanabe, T. (2010). Advances in visual perceptual learning and plasticity. *Nature Reviews Neuroscience*, 11(1), 53–60. https://doi.org/10.1038/nrn2737

Schoups, A., Vogels, R., Qian, N., & Orban, G. (2001, August 2). Practising orientation identification improves orientation coding in V1 neurons. *Nature*, 412, 549.

Seitz, A. R. (2017). Perceptual learning. *Current Biology*, 27(13), R631–R636.

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

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

Szpiro, S. F. A., & Carrasco, M. (2015). Exogenous attention enables perceptual learning. *Psychological Science*, 26(12), 1854–1862, https://doi.org/10.1177/0956797615598976.

Szpiro, S. F., Lee, Y. A., Wright, B., & Carrasco, M. (2013). Alternating training between tasks enables visual perceptual learning. *Journal of Vision*, 13(9): 249, https://doi.org/10.1167/13.9.249. [Abstract]

Taylor, M. M., & Creelman, C. D. (1967). PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America*, 41, 782–287.

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

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

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.

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. A., & Lu, Z.-L. (2018). Assessing the detailed time course of perceptual sensitivity change in perceptual learning. *Journal of Vision*, 18(10): 1068, https://doi.org/10.1167/18.10.1068. [Abstract]

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.