

Trial-dependent psychometric functions accounting for perceptual learning in 2-AFC discrimination tasks

Florian Kattner

Department of Psychology,
University of Wisconsin-Madison, Madison, WI, USA
Institute of Psychology, Technische Universität Darmstadt,
Darmstadt, Germany

Aaron Cochrane

Department of Psychology,
University of Wisconsin-Madison, Madison, WI, USA

C. Shawn Green

Department of Psychology,
University of Wisconsin-Madison, Madison, WI, USA

The majority of theoretical models of learning consider learning to be a continuous function of experience. However, most perceptual learning studies use thresholds estimated by fitting psychometric functions to independent blocks, sometimes then fitting a parametric function to these block-wise estimated thresholds. Critically, such approaches tend to violate the basic principle that learning is continuous through time (e.g., by aggregating trials into large “blocks” for analysis that each assume stationarity, then fitting learning functions to these aggregated blocks). To address this discrepancy between base theory and analysis practice, here we instead propose fitting a parametric function to thresholds from each individual trial. In particular, we implemented a dynamic psychometric function whose parameters were allowed to change continuously with each trial, thus parameterizing nonstationarity. We fit the resulting continuous time parametric model to data from two different perceptual learning tasks. In nearly every case, the quality of the fits derived from the continuous time parametric model outperformed the fits derived from a nonparametric approach wherein separate psychometric functions were fit to blocks of trials. Because such a continuous trial-dependent model of perceptual learning also offers a number of additional advantages (e.g., the ability to extrapolate beyond the observed data; the ability to estimate performance on individual critical trials), we suggest that this technique would be a useful addition to each psychophysicist’s analysis toolkit.

Introduction

One common assumption, instantiated in numerous theoretical models in the domains of psychology, neuroscience, and computer science, is that learning is a continuous function of experience. For example, this assumption underlies all models that use some form of a delta rule procedure (Casey & Sowden, 2012; Rumelhart, Hinton, & Williams, 1986; Spratling & Johnson, 2006). Here, in each learning epoch, the learner makes a prediction regarding the correct output, and then receives feedback as to the true correct output. The learner then computes the difference between their prediction and the true correct output and uses this to update the next prediction. When done repeatedly over time, this process will tend to gradually move the learner’s predictions into alignment with the true correct outputs. Learning is also modeled as a continuous process in many purely associative learning models (Bejjanki, Beck, Lu, & Pouget, 2011; Guenther, Ghosh, & Tourville, 2006; Rosenblatt, 1958; Spratling & Johnson, 2001). These models regularly use some form of Hebbian learning principle, wherein the strength of the connection between two nodes is updated after each learning epoch by an amount proportional to the extent to which the two nodes were simultaneously active during the learning epoch. Finally, Bayesian learning models are inherently continuous in nature, as each observed training example increases or decreases the probability that a particular estimate/hypothesis is correct (by an amount that depends on the strength of the evidence

Citation: 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, doi:10.1167/17.11.3.

doi: 10.1167/17.11.3

Received April 4, 2016; published September 6, 2017

ISSN 1534-7362 Copyright 2017 The Authors



provided in the training example and the prior probability that the particular estimate/hypothesis was correct; (Jacobs & Kruschke, 2011; Michel & Jacobs, 2007). Thus, although the general spirit as well as the detail-level implementation of these models may vary substantially, all instantiate this same basic principle that learning is a continuous process wherein small changes in ability accumulate via experience (Lu, Hua, Huang, Zhou, & Doshier, 2011; Mazur & Hastie, 1978; Petrov, Doshier, & Lu, 2005).

Not surprisingly, models in the field of perceptual learning share this same fundamental assumption that learning mechanisms are inherently incremental (Herzog & Fahle, 1998; Law & Gold, 2009; Lu et al., 2011; Petrov et al., 2005; Poggio, Fahle, & Edelman, 1992; Sotiropoulos, Seitz, & Seris, 2011; Vaina, Sundareswaran, & Harris, 1995; Zhaoping, Herzog, & Dayan, 2003). Indeed, theoretical models in this domain frequently have at their core one of the three broad types of learning rules/processes above (i.e., delta rule, associative/Hebbian, Bayesian). Interestingly, perceptual learning is even posited to be continuous in conditions that may not initially seem to easily support such learning. Take, for instance, the case of block feedback (Herzog & Fahle, 1997). In training tasks that employ block feedback, participants do not receive feedback regarding their accuracy after each trial. Instead, they are given their average accuracy across the entire previous block of trials after the block is finished. This poses a challenge for many of the models aforementioned, which require an explicit error signal to update behavior (i.e., the type of signal that would typically come from trial-by-trial feedback). However, one influential theoretical model in the field of perceptual learning produces continuous changes in performance even in this block feedback case. In this model, participants can use external and internal signals to update behavior. If, as is true in block feedback designs, there is not an external learning signal available to drive learning on each trial, the model will instead use internal estimates to alter performance continuously (with the internal signals being updated whenever external feedback is provided; Liu, Doshier, & Lu, 2014).

Thus, given the fact that essentially all theoretical models in the domain of perceptual learning suggest that learning should be continuous with experience, it is interesting to note that, in most behavioral experiments in this domain, learning is accounted for in a discontinuous manner. Rather than modeling changes in behavior as a continuous process using completely trial-dependent parameters, learning is often accounted for by first computing performance in discrete “blocks” of trials and then using the differences across those blocks (or fitting a parametric function to block performance as the measure of learning; Ball & Sekuler,

1987; Beard, Levi, & Reich, 1995; Doshier & Lu, 1998; Fahle & Edelman, 1993; Fahle & Morgan, 1996; Fendick & Westheimer, 1983; Gantz, Patel, Chung, & Harwerth, 2007; Seitz, Nanez, Holloway, Tsushima, & Watanabe, 2006; Yu, Klein, & Levi, 2004). For example, in one common method, the learning data is first subdivided into discrete blocks of trials, with the block size typically being based upon the experimental methods that were employed and ranging anywhere from 50–700 trials; (Ball & Sekuler, 1987; Fahle & Morgan, 1996). Then, a psychometric function is fit to the data within each block (e.g., logistic, Weibull, or cumulative Gaussian; Coates & Chung, 2014; Crist, Kapadia, Westheimer, & Gilbert, 1997; Fahle & Edelman, 1993) and a threshold value is calculated (e.g., 79% threshold). The difference in this threshold value from early blocks in training to late blocks in training is then used as the quantification of learning. In another common method, the threshold values for blocks are parametrically fit with a monotonically decreasing function (e.g., power, exponential; Astle, Blighe, Webb, & McGraw, 2015; Chung, 2011; Coates & Chung, 2014; Herzog & Fahle, 1997, 1999; Levi, Polat, & Hu, 1997; Matthews, Liu, Geesaman, & Qian, 1999; for a review, see Doshier & Lu, 2007).

Critically though, one important implicit assumption of such fitting procedures is that the parameters of the function are *not* changing over the block of trials being considered (e.g., the fitting in these cases necessarily assume that the data is generated by a constant level of performance). Thus, even when using parametric fits to block thresholds, performance is assumed to be stationary within each block and the most precise estimate of performance and learning is at the aggregated block level (alternatively, each block threshold must be taken to represent a particular trial in the block, e.g., the middle trial or the first trial). The process of fitting a learning function to block thresholds is problematic itself due to the errors inherent in sequentially modeling hierarchical data; see, e.g., Moscatelli, Mezzetti, & Lacquaniti, 2012.

This same implicit assumption regarding within-block stationarity of performance also underlies essentially all adaptive techniques for quickly estimating thresholds (e.g., staircases, PEST, QUEST, etc.; see Treutwein, 1995). Indeed, using an adaptive technique to estimate a threshold makes little sense if the threshold is actively changing during the estimation. Finally, this assumption of stationarity is present in any statistics that simply aggregate performance over an entire block without fitting. This includes analyses built upon signal detection theory (e.g., d' analysis assumes that the particular pattern of hits and false alarms across a block of trials is driven by a constant sensitivity) as well as any technique wherein performance is quantified as a simple average over blocks of

trials (e.g., percent correct). In essence, by using the aforementioned approaches, participants are being modeled as not changing at all *within* blocks of trials and instead are only free to improve *in between* blocks of trials (i.e., learning in a stepwise fashion). Because such a stepwise function is in direct contrast to our theoretical understanding of learning as a continuous function characterizing the relation between improvement and experience, in the present paper we present a new method of analyzing perceptual learning data to account for continuous changes in performance as a function of experience. Specifically, we employ a standard psychometric function whose parameters are allowed to change continuously through time. This is conceptually identical to fitting a psychometric function to all data points as a single block, but parameterizing nonstationarity rather than assuming within-block stationarity (Fründ, Haenel, & Wichmann, 2011). By fitting the psychometric function to the largest possible “block” (i.e., all trials) we reduce noise introduced by factors other than perceptual ability, and by estimating learning as a function of the smallest possible “block” (i.e., each individual trial) our estimates better reflect the continuous nature of learning. In addition, we improve upon parametric fits to block estimates by requiring fewer free parameters, while also providing simultaneous fits to stimulus (threshold) and time (learning) dimensions.

Here, we show that our continuous time-parametric model provides a better fit, without overfitting, to perceptual learning data than the more traditional trial-independent, nonparametric approach of fitting psychometric functions to data in consecutive blocks of trials. This is perhaps not surprising given that block models necessarily take a functional form that is inconsistent with our beliefs about actual human learning. Furthermore, in addition to simply providing a better fit to perceptual learning data, the continuous time-parametric model also offers a number of other empirical (e.g., more accurate extrapolation of performance) and theoretical advantages (e.g., ability to use all data in assessing the functional form of learning; provides a natural method for estimating certain important trials, such as the first and last trial of training) over standard nonparametric block fitting or parametric fits to blocks. We therefore suggest it will be a valuable addition to every psychophysicist’s toolkit.

Method

Perceptual learning data/tasks

Data from two different standard perceptual learning tasks was used in the current analysis. Both data

sets, including one examining orientation discrimination training ($N = 7$) and one examining stereoacuity training ($N = 7$), overlap with previously published data sets (Green, Kattner, Siegel, Kersten, & Schrater, 2015; Snell, Kattner, Rokers, & Green, 2015). In the following material, we briefly describe the basic training methods for the data that is considered (Note: For each of the following tasks, participants underwent brief pretests without feedback prior to training on both the to-be-trained task as well as various transfer measures; however, because the focus of the current manuscript is on fitting learning curves this data is not considered).

Orientation discrimination training methods

For full task methods, see Green et al. (2015). Briefly, in the orientation discrimination training task, on each trial, participants were presented with a central “T” (either upright or upside down) as well as a full-contrast Gabor patch at an eccentricity of 10° below the “T.” The orientation of the Gabor was drawn from a uniform random distribution between 30° and 60° . After the stimuli were presented, the participants were required to first respond to the orientation of the “T” (by pressing the “w” or “s” key for upright or upside down, respectively), and then to the orientation offset of the Gabor relative to 45° (by pressing the right or left arrow key for clockwise or counterclockwise, respectively). Participants completed 3,800 such trials, distributed over four different days.

Stereoacuity training methods

For full task methods, see Snell et al. (2015). In the stereoacuity training task, on each trial, two white three-dimensional rectangles, offset relative to one another in depth, were presented. The size of the offset was drawn from a uniform distribution between 0 and 60 arcsec. The participants’ task was to indicate which square appeared closer in depth. Participants completed 7,500 of such trials, distributed across five different days.

Parametric model of continuous perceptual learning

Because both tasks involved participants making two-alternative forced choice (2-AFC) decisions on stimuli that varied in signal intensity, the continuous model was built upon a generalized psychometric function (Equation 1), relating an observer’s responses to stimulus intensity x (e.g., orientation or stereo offsets; see Supplemental Materials for additional fitting details related to software, etc.):

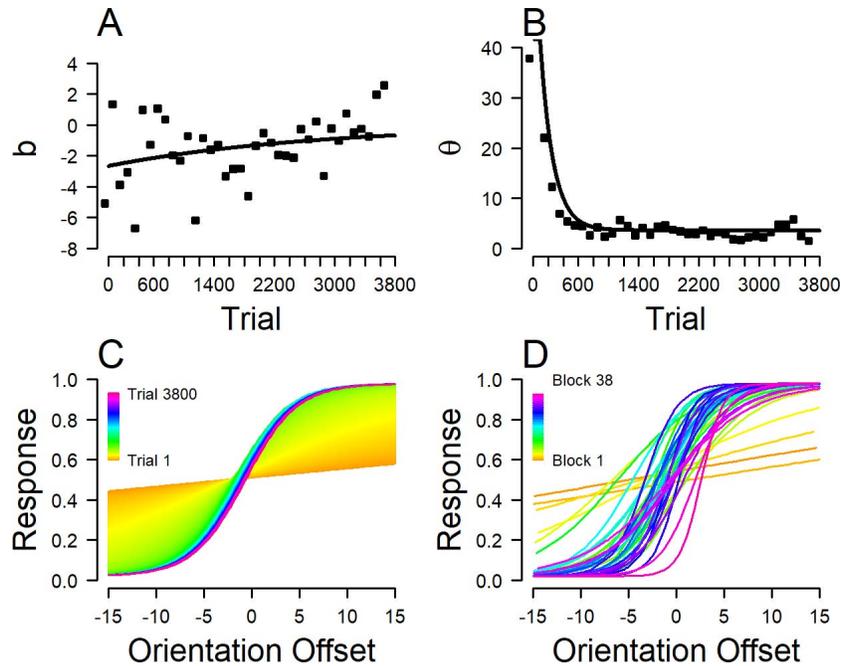


Figure 1. Illustration of the parameters b and θ as fit using the block and continuous models of psychophysical data with an exemplar orientation discrimination subject. (A) Change of the bias value b fit within 38 independent successive blocks (black squares) or as a continuous function of trial number (solid line). (B) Change of the threshold value θ fit as 38 independent successive blocks (black squares) and as a continuous function of trial number (solid line). (C) The resulting trial-dependent psychometric function as estimated with the continuous model (0 = counterclockwise, 1 = clockwise). As is clear, the continuous approach models performance as changing smoothly through time. (D) The resulting psychometric functions in 38 independent 100-trial blocks of training (0 = counterclockwise, 1 = clockwise). This approach is, in essence, only allowing for changes in performance across blocks of trials.

$$f(x; b, \theta) = \lambda + \frac{1 - 2\lambda}{1 + k^{\frac{b-x}{\theta}}} \quad (1)$$

Equation 1 contains two constants: The parameter k was defined as 3.76 ($= 0.79/0.21$) in order to estimate 79% thresholds. λ refers to a lapse parameter. In essence, this can be considered the probability that a participant will show a total “lapse” of attention in which case his/her performance will be unrelated to the stimulus that was presented; in practice, this is used to account for trials at very high levels of stimulus strength that the participant nonetheless answers incorrectly (see Klein, 2001). This lapse value was held constant at 0.02. (Note that values for the lapse parameters between 0 and 0.1 were assessed—for the most part, the particular value did not affect the quality of fits, although there was a tendency for the largest values to somewhat degrade the fits.) The two remaining parameters b and θ refer to the bias and threshold of the psychometric function, respectively. To account for continuous perceptual learning, these two parameters (b and θ) were themselves fit as functions of time (t ; see Equations 2 and 3 as follows). Because the focus of this article is not on the exact functional form of those parameters in relation to time

(we note that identifying the exact functional form of the change function is beyond the scope of this paper; see Discussion and Supplemental Materials, and Kattner, Cochrane, Cox, Gorman, & Green, 2017 for an alternative parameterization), we chose to model bias as a two-parameter exponential function of time (Equation 2) and threshold as a three-parameter exponential function of time (Equation 3).

$$b(t) = b_0 \cdot e^{-\frac{t}{b_1}} \quad (2)$$

$$\theta(t) = pre - (pre - post)e^{-\frac{t}{\tau}} \quad (3)$$

Continuous perceptual learning in the two sets of data can thus be accounted for by a psychometric function with two constants (k and λ) and five independent parameters, with b_0 referring to initial bias, pre to the initial threshold, $post$ to the final asymptote of the threshold, and two slope parameters for bias and threshold (b_1 and τ , respectively).

The relationship between trial-dependent parameter estimates (bias and threshold) and the time-evolving psychometric function is illustrated in Figure 1 (panels A, B, C) for an exemplar participant trained on the orientation discrimination task.

Nonparametric (block) model of perceptual learning

As a standard against which the continuous time parametric model could be compared, we also fit the data via a block-based method commonly used (Doshier & Lu, 2000) in the perceptual learning field (see Figure 1, panel D). For each participant and task, the data was first divided into blocks of 100 trials each. A single logistic function (Equation 1; two free parameters) was then fit to each block. The relationship between block-by-block parameter estimates (bias and threshold) and the resulting discrete psychometric functions in each block are illustrated in Figure 1 (panels A, B, and D) for an exemplar orientation discrimination participant.

Hybrid model (see Supplemental Materials)

Our main interest in this paper is in comparing the two analysis approaches described already (approach #1: the nonparametric block model where thresholds are fit to blocks of aggregated data thus not assuming a trial-dependent change in parameters, e.g., Crist et al., 1997; Fahle & Edelman, 1993; approach #2: the continuous time parametric model where thresholds are fit by considering trial-by-trial changes in the parameters of the psychometric function). However, it is worth noting that there is a third approach that is, to some extent, an intermediate between a fully time-continuous model and a fully block model. Namely, it is possible to account for perceptual learning by first aggregating data within discrete blocks of trials and then fitting a continuous model to these “block-averaged” response probabilities. Because parametric fitting to block thresholds is a common approach to analyzing data in the literature (e.g., Chung, 2011; Coates & Chung, 2014; Fründ et al., 2011; see Doshier & Lu, 2007 for a review of earlier studies) we present the results of this type of “hybrid” model in the Supplemental Materials along with several other alternative models, all of which use the same learning functions (Equations 2 and 3; see Discussion).

Comparing continuous parametric and block nonparametric analyses

In comparing the continuous parametric and block-based nonparametric analysis approaches, we examined several basic aspects of model quality. The first was simply to examine how well the model captures the full pattern of data. To this end, after fitting both the continuous time and the block model to the full training data for each participant, several measures were assessed, including Akaike and Bayesian infor-

mation criteria (AIC and BIC, respectively). Both metrics provide estimates of the quality of a model relative to other models. More specifically, both involve a calculation of the likelihood function (i.e., probability of the data given the model; see Equation 4, with r being the observed binary responses) that is then penalized based upon the number of parameters in the model (penalty term = $[-2\log L + kp]$, with L being the likelihood function, p being the number of parameters in the model, and k being an additional penalty that differs for AIC and BIC). In particular, with large numbers of parameters, this penalty is greater in the case of BIC than AIC (the term k is set to 2 for AIC and $\log(p)$ for BIC). In addition to AIC and BIC, we also calculated χ^2 measures (accounting for the discrepancy between theoretical/fitted and observed data, see Supplementary Materials for equation; cf. Klein, 2001).

$$\log L = (\sum \log(f(x; b, \theta)))r + (1 - \log(f(x; b, \theta)))(1 - r) \quad (4)$$

Second, one major concern in essentially all data modeling is related to overfitting—when the model captures random fluctuations/noise in the data rather than only capturing true signal (i.e., in the case of perceptual learning, the true signal would be actual changes in performance). Overfitting becomes a greater and greater concern as models increase in complexity (e.g., increases in the number of free parameters; see Figure 2). To account for overfitting, the quality of each model fit was assessed in a standard train/test procedure. Specifically, the models were first fit to data on the odd trials only (i.e., 1,900 and 3,750 orientation and stereo discrimination trials, respectively). The quality of the resulting model fit was then assessed with respect to data on the even trials by calculating AIC, BIC, and χ^2 . The rationale here is that the training set (i.e., the odd trials) should be generated from the same basic perceptual sensitivity as the test set (i.e., the even trials). Thus, by examining how well the fits derived from the training set (odd trials) match the data in the test set (even trials), we can assess the extent to which the two competing models properly fit the data, without overfitting the data.

Third, 95% confidence intervals around the fits were computed via bootstrapping (i.e., drawing random samples with replacement from the individual trial sequences, fitting psychometric functions and calculating threshold of each sample, and determining the confidence intervals from the 97.5% and 2.5% percentiles; see Supplemental Materials for additional detail on the bootstrapping procedure). The width of the confidence intervals was then contrasted between the models. Like already mentioned, this provides an estimate of model quality, again, in particular the

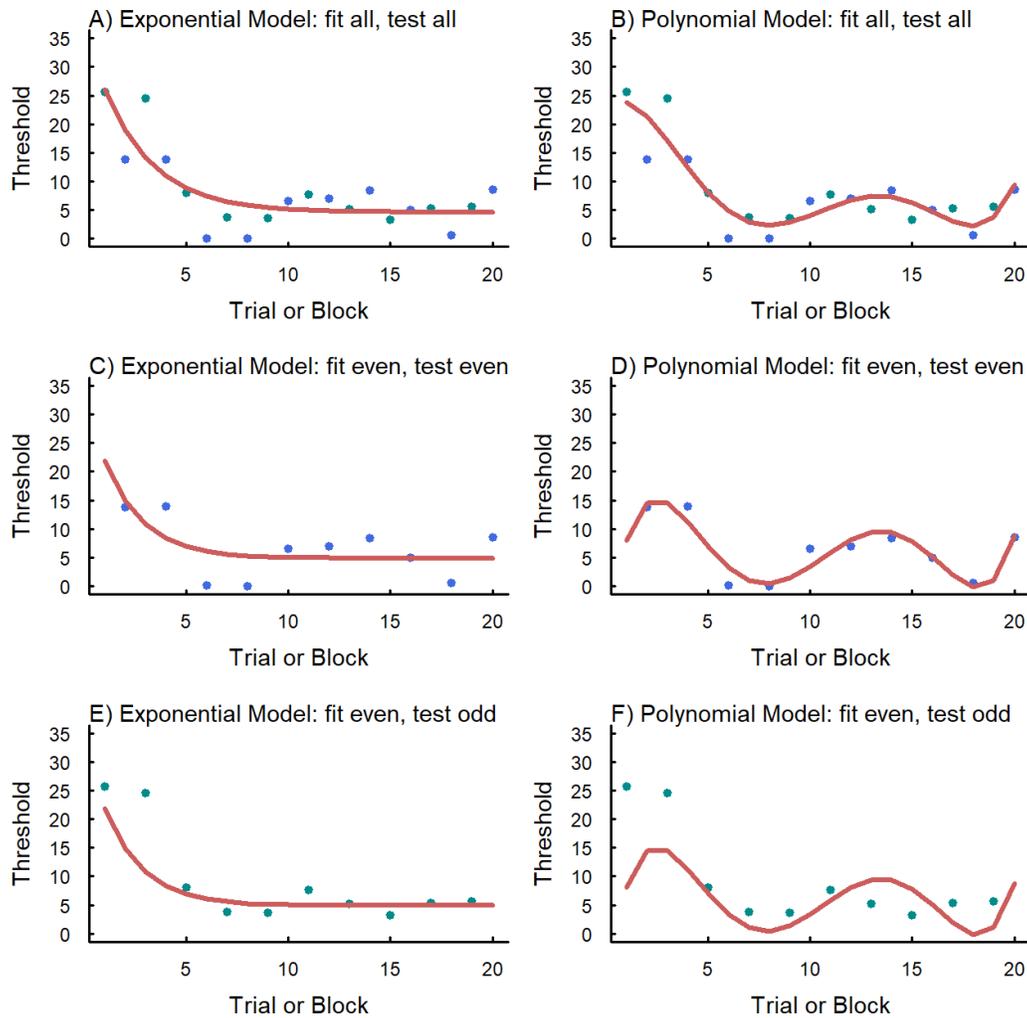


Figure 2. Illustration of how overfitting can be detected. (A) Fit to an arbitrary time series data using a less complex model (exponential). (B) Fit to the same arbitrary time series data as in A, but using a more complex model (high-order polynomial). In examining how well the two models fit the data in A and B, it is clear that both do a reasonable job of predicting the position of the data points with the more complex model, if anything, doing a better job. (C) Fit to just the even trials of the time series data using the less complex model. (D) Fit to just the even trials of the time series data using the more complex model. Again, both models do a reasonable of predicting the position of the data points. (E) The fit on the even trials from the less complex model continues to do a good job of predicting the position of the untrained (odd) trials. (F) The fit on the even trials from the more complex model does a quite poor job of predicting the position of the untrained (odd) trials. This is the hallmark of an overfitting model. It does a good job of making predictions about the trained data set, but performance is markedly poorer for the untrained data.

extent to which the model is susceptible to overfitting. A model that is more prone to overfitting might produce markedly different estimates depending on the particular (perhaps idiosyncratic) set of trials that is considered. A model that is less prone to overfitting should produce essentially the same estimates regardless of the exact set of trials that is considered.

Fourth, both the models were fit only to an initial portion of the data and then used to predict fits of the remaining trials. This tests a final critical aspect of perceptual learning data analyses – the ability to make a forward prediction. Specifically, thresholds were fit to the responses on the initial 1,500 orientation and 3,000 stereo discrimination trials, respectively, and the fitted

thresholds were then extrapolated over the remaining trials of the respective task. The extrapolated estimates were then contrasted against the true data by calculating the AIC, BIC, and χ^2 measures. Note that the nonparametric block approach provides no natural method of extrapolation (as it does not implement any particular functional form). Thus, for the block model, the extrapolation was obtained by fitting continuous functions (Equations 2 and 3) to the threshold and bias parameter estimates obtained for each block (i.e., with constant biases and thresholds for all trials within a block) of the initial subset of the data. These parameter functions can then be used to predict threshold for the remaining blocks.

Subject	Continuous model				Block model			
	<i>logL</i>	<i>AIC</i>	<i>BIC</i>	χ^2	<i>logL</i>	<i>AIC</i>	<i>BIC</i>	χ^2
O1	–1477	2964	2995*	3998*	–1403*	2957*	3155	4094
O2	–1643	3296	3327*	3723	–1508*	3169*	3367	3388*
O3	–1626	3263	3294*	3829	–1540*	3232*	3430	3652*
O4	–1086	2182*	2213*	3415	–1029*	2210	2408	3147*
O5	–2380	4770	4801*	3810*	–2268*	4687*	4885	3898
O6	–1281	2571	2603*	4335*	–1198*	2549*	2747	4413
O7	–1671	3351	3383*	3747*	–1532*	3215*	3413	4102
S1	–3019	6049	6083*	7929	–2838*	5975*	6366	7430*
S2	–3972	7955	7989*	7638	–3825*	7949*	8340	7534*
S3	–2357	4724	4758*	9201	–2157*	4613*	5004	7881*
S4	–4271	8551*	8586*	7713*	–4141*	8582	8973	7766
S5	–2808	5626	5661*	7815	–2650*	5600*	5991	7780*
S6	–4369	8748*	8783*	7564	–4231*	8762	9153	7562*
S7	–3590	7191	7225*	7732	–3423*	7146*	7537	7499*

Table 1. Overall analysis of model fits: Goodness of fit metrics (*log* likelihoods, *AIC*, *BIC*, and χ^2) for the continuous model ($np = 5$) and block model ($np = 76$ and $np = 150$, respectively) fit to $nd = 3,800$ orientation discrimination trials (subjects O1–O7) and $nd = 7,500$ stereo discrimination trials (subjects S1–S7). Asterisks indicate the model with the best relative fit to the data (note that this does not indicate “statistical significance” in the traditional null-hypothesis sense).

Results

Overall fit

The models were first fit to responses in all 3,800 and 7,500 trials of orientation and stereo discrimination training, respectively. The resulting goodness of fit metrics are summarized in Table 1 (see the Supplemental Materials for information on the hybrid models). Both models clearly fit the data well overall. Of interest is that the block model tends to show better performance as determined by *AIC* values (better for 11 out of 14 participants), whereas the continuous model shows better performance as determined by *BIC* values (better for 14 out of 14 participants). Less consistent support for either model has been obtained with the χ^2 metric (i.e., the discrepancy between observed and predicted data was smaller for the continuous model in five out of 14 participants). Given that the major difference between *AIC* and *BIC* is the extent to which greater model complexity is penalized, it was next of interest to examine whether the better performance seen in *AIC* values in the block model is due to overfitting.

Test for overfitting: Train/test analysis

For both tasks, the goodness of fit of the models was evaluated by fitting the models to the responses on odd trials and testing with regard to the responses on even trials. Table 2 shows the resulting *AIC*, *BIC*, and χ^2

metrics for the block and continuous models fit to the data of each participant (see Supplemental Materials for the hybrid models). As can be seen in Table 2, the test data was consistently fit better by the continuous model (providing a markedly better fit for all 14 participants regardless of the measure of model fit employed). It is remarkable, given that the continuous model only has a fraction of the number of free parameters that the block model has, that this more-parsimonious fitting method is clearly a better fit to the test data. This pattern of results suggests very strongly that while the nonparametric block approach appeared to do a reasonable job when examining the fit to the full data set, it was very likely in fact dramatically overfitting the data (i.e., block threshold estimates are likely fitting some noise). Meanwhile, the continuous model, which is (a) far less complex in terms of number of parameters than the block model and (b) instantiates a strong belief about the manner in which the data should be generated, appears to be considerably less susceptible to overfitting. Just as larger blocks will reflect perceptual ability more accurately by averaging over more noise in the data, fitting all of the data as one nonstationary block (i.e., the continuous parametric fit) minimizes the influences of noisy data on threshold estimates.

Learning gains and confidence intervals

The amount of perceptual learning in both discrimination tasks can be quantified (with both models) by subtracting the final threshold estimates from the initial threshold estimate (e.g., first 100 trials minus

Subject	Continuous model				Block model			
	<i>logL</i>	<i>AIC</i>	<i>BIC</i>	χ^2	<i>logL</i>	<i>AIC</i>	<i>BIC</i>	χ^2
O1	−792*	1594*	1621*	2201*	−882	1916	2061	4459
O2	−820	1651*	1678*	1852*	−814*	1780	1926	2481
O3	−785*	1580*	1607*	1802*	−807	1766	1911	2313
O4	−520*	1050*	1078*	1555*	−564	1279	1425	2405
O5	−1189	2387*	2415*	1900*	−1181*	2514	2660	2345
O6	−628*	1267*	1294*	2175*	−672	1496	1642	3341
O7	−841*	1692*	1720*	1892*	−866	1885	2030	3328
S1	−1496*	3002*	3033*	3858*	−1578	3455	3742	5923
S2	−2008*	4026*	4057*	3855*	−2084	4468	4754	5105
S3	−1205*	2419*	2450*	4819*	−1243	2786	3073	5995
S4	−2157*	4324*	4355*	3869*	−2226	4752	5039	4939
S5	−1365*	2739*	2770*	3753*	−1413	3127	3413	4669
S6	−2184*	4379*	4410*	3797*	−2264	4827	5114	4738
S7	−1788*	3586*	3617*	3788*	−1836	3972	4259	5042

Table 2. Overfitting analysis for the orientation (subjects O1–O7) and stereo (subjects S1–S7) discrimination data. The continuous ($np = 5$) and the block model ($np = 76$ and $np = 150$) were fit to $nd = 1,900$ and $nd = 3,750$ even orientation and stereo discrimination trials, respectively. Models were then tested with regard to the same number of odd trials. Asterisks indicate the model with the best relative fit to the data.

last 100 trials; note that the initial orientation and stereo discrimination threshold estimates were constrained to a maximum of 90° and 1.5 arcmin, respectively). The overall amount of learning estimated with the two models was almost identical: For the orientation discrimination task, the block model found threshold improvements of $M = 19.34^\circ$ ($SD = 23.67$), whereas the continuous model estimates an average reduction in threshold of $M = 19.28^\circ$ ($SD = 20.94$). For the stereo discrimination task, improvements of 25.07 arcsec ($SD = 29.49$) were found with the block model, whereas the continuous model suggests a threshold decrease of 25.88 arcsec ($SD = 22.16$). The estimated improvements did not differ significantly between models for either task, $p = 0.71$ and $p = 0.90$, respectively (using nonparametric Wilcoxon rank tests; similar nonsignificant results are found using t tests).

However, differences were found with regard to the confidence of the fitted thresholds that were reached as a result of perceptual learning. The individual 79% thresholds and 95% confidence intervals, as obtained with the continuous and with the block model, are illustrated in Figures 3 and 4 for the orientation and stereo discrimination task, respectively. For the orientation discrimination data, a one-sample Wilcoxon signed-rank test revealed that the confidence intervals orientation discrimination thresholds estimated for the final 100-trial block were significantly smaller with the continuous model ($M = 1.49^\circ$; $SD = 1.27^\circ$) than with the block model ($M = 4.59^\circ$; $SD = 3.27^\circ$), $p = 0.03$ (see Figure 3). Likewise, the confidence intervals of the estimated stereo discrimination thresholds in the last

block were significantly lower with the continuous model ($M = 7.51$ arcsec; $SD = 7.96$ arcsec) than with the block model ($M = 25.93$ arcsec; $SD = 20.09$ arcsec), $p = 0.02$ (see Figure 4).

For both the continuous and the block model, the average CI in the last block was subtracted from the average CI in the first block for each data set in order to quantify how well both models capture learning-related decreases in uncertainty of the two models. With the continuous model, the median CI decrement was 71.80° for the orientation discrimination task, and 17.11 arcsec for the stereo discrimination data. In contrast, with the block model, the CI decrements were 15.56° and 11.67 arcsec for the orientation and stereo discrimination tasks, respectively.

Extrapolation analysis

The qualities of the continuous and the block models were further evaluated by fitting both models only to an initial portion of the training data (i.e., 1,500 and 3,000 trials for the orientation and stereo discrimination tasks, respectively), and then extrapolating the thresholds for the remaining trials of each training task, based on the fitted models. For the majority of data from both psychophysical tasks, the thresholds extrapolated based on the continuous model fit the actual data better than did the thresholds extrapolated based on the block model. The exact goodness-of-fit measures for the block and continuous models extrapolated to data from the two tasks are summarized in Table 3 (see the Supplemental Materials for the hybrid model). As

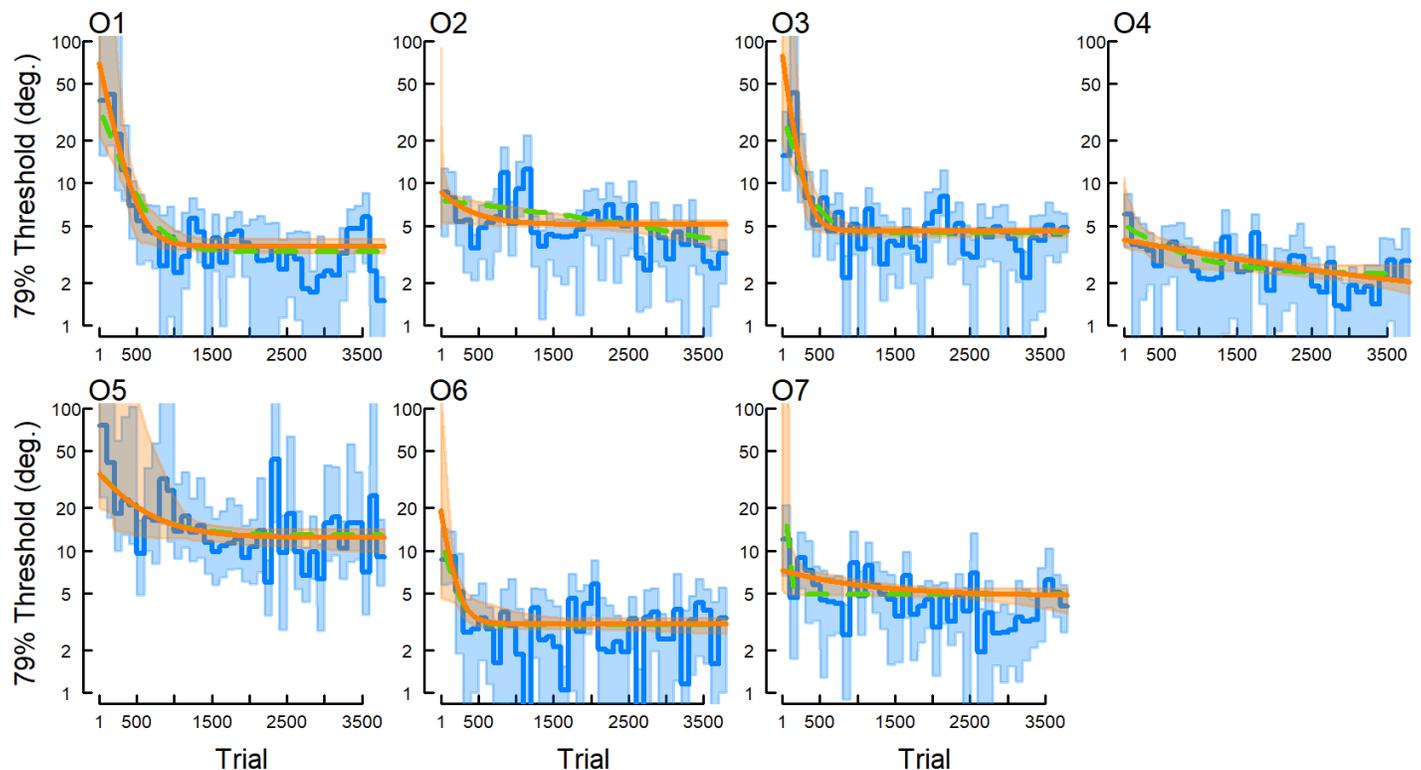


Figure 3. Individual orientation discrimination thresholds based on the 38 separate logistic fits (Equation 1) to 100-trial blocks (block model; blue lines), and the five-parameter continuous model (orange lines). The shaded areas represent the respective bootstrapped 95% confidence intervals for the block and continuous fits, respectively. The dashed green line refers to the hybrid model thresholds (see Supplemental Materials).

expected, for six out of seven orientation discrimination participants and for all stereo discrimination participants, the BIC metrics of extrapolated psychometric functions reached a better fit with the continuous model than with the block model. Only for one participant (O6), responses on the first 1,500 trials could be extrapolated better with the block model (based on BIC), probably due to a slower rate of learning (i.e., the continuous model may not have identified a consistent decrease in threshold during this period; see Figure 3).

Discussion

The majority of theoretical models across many domains of psychology, including the domain of perceptual learning, consider learning to be a process that occurs continuously with experience. However, despite this, most empirical studies in this domain have modeled learning as arising via a discontinuous process. Indeed, participant data in this field is nearly always first separated into distinct blocks of trials for analysis, with block sizes typically being guided by experimental decisions (e.g., how many trials are

feasible per day). From there, whether the analyses involve data fitting (e.g., fitting performance across each block with a psychometric function) or simple aggregation/computation (e.g., percent correct across the block; d' across the block), all share the implicit assumption that there is no significant change in performance *within* blocks and instead only allow for learning to occur *in-between* blocks.

Given the clear mismatch between the theoretical and analytical approaches in this domain, here we sought to develop a method to bring these approaches into better alignment. Specifically, using data collected from two perceptual learning experiments, we compared two main analytical data fitting methods—the standard nonparametric approach (fitting psychometric functions to blocks of trials) and a new continuous time parametric approach that allows for a trial-dependent continuous change in the parameters of the psychometric function. Consistent with existing theory in the field, the continuous time parametric model of perceptual learning provided a more parsimonious account for the data than the standard nonparametric trial-independent approach. Importantly, our new continuous time parametric model did not do so by producing totally different estimates than the block-based approach. Rather, the fact that the core estimates

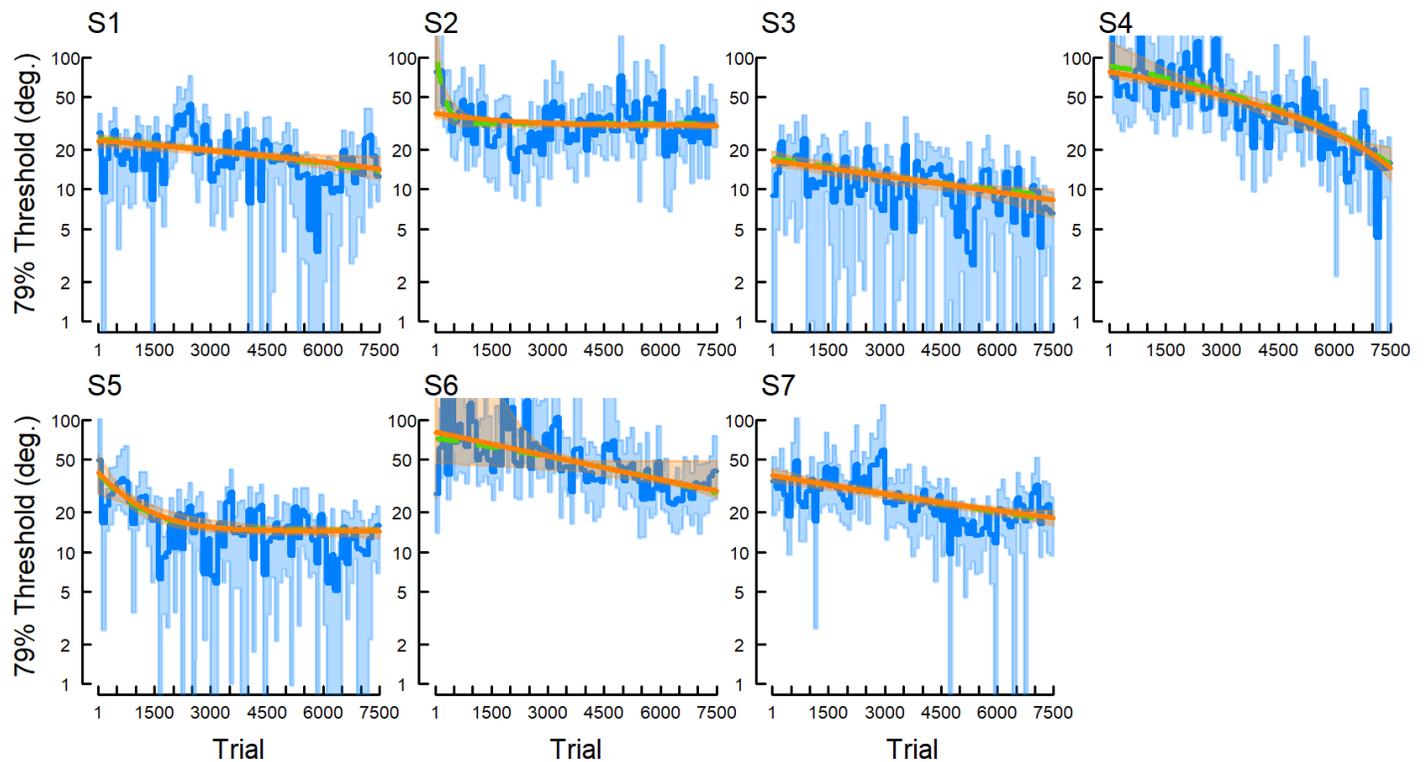


Figure 4. Individual stereo discrimination thresholds based on the 75 separate logistic fits (Equation 1) to 100-trial blocks (block model; blue lines), and the five-parameter continuous model (orange lines). The shaded areas (blue and orange for the respective models) represent bootstrapped 95% confidence intervals for the respective fits. The dashed green line refers to the hybrid model thresholds (see Supplemental Materials).

of thresholds/learning produced by the standard methods were quite similar to those produced by our new approach speaks to the validity of the new approach.

Overfitting

Perhaps the largest difference between the continuous approach and the nonparametric block approach

Subject	Continuous model				Block model			
	<i>logL</i>	<i>AIC</i>	<i>BIC</i>	χ^2	<i>logL</i>	<i>AIC</i>	<i>BIC</i>	χ^2
O1	-788*	1587*	1616*	2015*	-871	1802	1974	5350
O2	-991*	1993*	2021*	1866*	-1104	2267	2439	2570
O3	-918*	1847*	1875*	2127*	-1055	2170	2342	5876
O4	-599*	1207*	1236*	1511*	-2286	4632	4804	25296
O5	-1417*	2845*	2874*	2149*	-1467	2994	3167	2371
O6	-874	1759	1788	6618	-770*	1601*	1773*	4142*
O7	-970*	1950*	1979*	2113*	-1017	2094	2266	2668
S1	-1871*	3752*	3784*	2902*	-3118	6357	6741	4556
S2	-2477*	4965*	4997*	6152*	-2506	5132	5516	6745
S3	-1389*	2788*	2820*	8401*	-1430	2981	3365	9941
S4	-2542	5094	5126*	3713*	-2474*	5068*	5452	3771
S5	-1566*	3142*	3174*	5112*	-1656	3431	3816	8440
S6	-2683	5377	5409*	4020*	-2573*	5265*	5650	4176
S7	-2081*	4172*	4204*	3258*	-3111	6342	6727	4542

Table 3. Extrapolation analysis: The continuous ($np = 5$) and the block model ($np = 20$ and $np = 50$) were fit to early training trials ($nd = 1,500$ and $nd = 3,000$ of orientation and stereo tasks, respectively), and thresholds were extrapolated by fitting the models to the remaining $nd = 2,300$ and $nd = 4,500$ orientation and stereo discrimination trials, respectively. Asterisks indicate the model with the better relative fit to the data (based on AIC, BIC, and χ^2).

was in the ability to fit the data without overfitting the data. Indeed, the large number of free parameters resulting from fitting multiple psychometric functions to a large number of blocks makes the nonparametric approach extremely flexible. This flexibility in turn means that the approach will tend to capture any and all fluctuations in the data (we note that this remains true of approaches that fit parametric functions to block threshold estimates; each threshold estimate being fit by the parametric function itself comes from a model fit—it is not in fact “raw data”—and thus such a model still has an extremely large number of free parameters and is exceptionally flexible, though less so than the nonparametric version; see Supplemental Materials for Hybrid model results). This includes fluctuations that arise due to noise alone, which can be substantial in 2-AFC experiments given the standard deviation associated with the binomial distribution. It also includes any number of non-monotonic fluctuations in performance that are unrelated to actual changes in perceptual ability (e.g., mind-wandering/failures of sustained attention, etc.). Data points consistent with these issues can be clearly seen just by simple examination of the block-by-block fits of both experiments, where threshold estimates in contiguous blocks commonly differed by substantial margins (i.e., in a manner that could not possibly be due to a true change in perceptual ability). The parametric continuous model, meanwhile, treats these occasional deviations as noise, and thus their presence did not substantially affect the estimate of participants’ true behavioral abilities at the given time points. Although overfitting is a concern any time data is modeled, it is of particular relevance to the domain of perceptual learning. Consider, for example, the simple question of “How much did participants improve?” at a given task. Because the nonparametric block approach tends to overfit noise, this will reduce confidence in early and final performance estimates and thus will reduce confidence in the estimated change between early and final performance (i.e., a few idiosyncratic points in either the first or last block may produce extremely different estimates of total learning).

Trial-specific estimation

In addition to the fact that the continuous time model simply provides a more trustworthy estimate of participant performance and learning, there are a number of other aspects of this method of fitting that may be useful to the field going forward. For instance, one benefit of the continuous parametric model is that it is capable of providing estimates of performance on particular trials of interest. Specifically, in learning

experiments, the most critical trials very commonly correspond with the first and last trials of training. Assuming that learning has roughly reached an asymptote by the end of training, the “last trial” estimates given by the continuous parametric approach and nonparametric block approach will tend to be quite similar (excluding issues related to overfitting noise in the block approach). This is because, once participants have hit a rough asymptote, their behavior closely approximates the key block-based assumption that performance is not changing substantially within a block. In contrast though, substantial differences between the nonparametric and parametric approaches are possible in their estimates of early performance. Indeed, given an exponential or power functional form, the early portion of training is when performance changes most rapidly from trial to trial (Badiru, 1992; Doshier & Lu, 2007; Heathcote, Brown, & Mewhort, 2000). Because the nonparametric model aggregates across the first 100 or more trials (Crist et al., 1997; Fahle & Morgan, 1996; Gantz et al., 2007; Z. Liu & Weinshall, 2000; Seitz et al., 2006); with 150, 80, 200, 2,000, and 640 trials, respectively), it is necessarily the case that this approach will collapse over a substantial amount of learning (i.e., the average performance across the block will be substantially better than the performance on the first few trials). Therefore, without using trial-dependent changes in the parameters of the psychometric function, the true “total” amount of learning that has occurred from the first trial of training to the last trial of training will necessarily be underestimated.

Testing functional form of learning

This issue also speaks to another benefit of the continuous parametric model—the ability to more fully examine the functional form of learning. Although our approach here was purely descriptive (i.e., the eventual functional form we chose was simply the one that provided the best overall fits to the data—see Supplemental Materials for other parameterizations; also see Kattner et al., 2017; Snell et al., 2015), the general framework can be easily used to test explicit predictions about the best fitting functional form. This is critical as the observed functional form of learning in a task limits the possible mechanistic models that need to be considered. As such, the question of functional form has thus been investigated in many different fields (Badiru, 1992; Heathcote et al., 2000; Newell & Rosenbloom, 1981) including the field of perceptual learning. For instance, it was reported that the improvements of Vernier acuity thresholds found in adults with amblyopia (not considering slope) across successive blocks (which

were distributed across several days) can be fit by an exponential function (Levi et al., 1997). Similarly, negative exponential functions were used to fit improvements in reading speed in visually impaired patients (Chung, 2011). Other researchers have meanwhile separately modeled improvements in thresholds and slopes (or the width of the psychometric function; Coates & Chung, 2014; Fründ et al., 2011). In seminal work by Doshier and Lu (2007), several different functional learning forms were contrasted, in particular, power versus exponential. The authors found that an exponential functional learning form provided the best fit to individual data (with power only fitting better for the aggregate across participants). Critically though, the authors in this case employed a staircase technique to train participants (140 trials per block). As noted already, because participants are likely to be learning rapidly during this first block, their performance at the end of the block (i.e., which is disproportionately what a staircase analysis focuses on) is almost certainly better than their performance at the beginning of the block. Thus, one likely outcome of this type of trial aggregation approach is to flatten the shape of the learning curve (i.e., by overestimating initial performance), which could in turn potentially affect the best fitting function. Although our experiments were not designed to speak to the exact issue of functional form (e.g., because the participants in our experiments underwent pretesting prior to training, which could also alter estimates of the functional form of learning), the overarching approach could easily be used to examine this question more closely.

The approach can also be extended to address related questions, such as whether the first 100 or 200 trials *should* be included in the full learning curve analysis or whether these trials should be thrown out/treated as practice trials as is common in the literature (i.e., whether there is an early stage of learning that is qualitatively and quantitatively different from the remainder of the learning process). Similarly, the parameterization can be extended or altered to determine whether there are changes in other aspects of the underlying performance functions such as the temporal characteristics of the response biases [e.g., (a) whether it necessary to allow the response bias to change over time or can it be set as a constant—see Supplemental Materials; (b) whether a separate function for the response bias be set for each “day” of the experiment] or the lapse/guess rate (e.g., in our case we assumed a constant lapse rate, but this could also change with training; for examples, see Fründ et al., 2011; Jones, Moore, & Amitay, 2015; Petrov, Doshier, & Lu, 2006) or in the best form of the probability distribution.

Estimating learning and transfer

Further, because the continuous time parametric-model fit has the potential to provide estimates of first trial and last trial performance, it could be additionally useful in designs seeking to address questions regarding total learning (i.e., by comparing first trial and last trial estimates rather than first block and last block), rate of learning, as well as questions related to transfer of learning (Kattner et al., 2017). In the latter cases, “transfer” could be calculated as the difference in performance on the final trial of the training task and the first trial of the transfer task, or the increase in learning rate in the transfer task when compared with the training task. It may also be possible in this endeavor to take advantage of the fact that the parametric model provides for a natural method to extrapolate beyond the data set (i.e., to estimate how performance would have continued to evolve if the participant had carried on with the training task, as compared to how they did perform when asked to perform a new transfer task). Such an extrapolation has no analogue in a nonparametric model that relies on aggregating across trials into blocks, in which performance on the last training block is frequently contrasted with performance on transfer block performance (Liu & Weinshall, 2000).

Finally, we note that the data of primary interest in many perceptual learning papers is not actually the training data, but is instead performance on pre- and post-tests (e.g., to determine whether there are improvements on some untrained task from pretest to posttest). The approach we have outlined here plays an important role in this type of design as well. Learning generalization (i.e., an improvement on an untrained task after training), can take multiple functional forms. There can be an immediate enhancement on the untrained task at posttest (i.e., performance on the first trial of the posttest exceeds performance on the last trial of the pretest). There can also be a change in the rate at which performance improves on the posttest (i.e., performance on the first trial of the posttest matches performance on the last trial of the pretest, but then performance on the post-test rapidly improves). We have referred to these different functional forms of generalization as “transfer” and “learning to learn” respectively (Kattner et al., 2017). Critically, it is easy to misidentify the functional form if a block approach is used to analyze pretest and posttest data (since such an approach does not given an estimate of immediate performance on the posttest nor an estimate of how performance changes throughout the posttest, but instead considers performance during the posttest to be stationary).

Limitations

Although, as we have shown here, modeling learning as a continuous function of experience on a task provides definite benefits in comparison to standard block-by-block analysis of perceptual learning, there are clear limitations to the approach as well. Many of these limitations are related to the strict functional form imposed by the continuous learning function, which would be an issue in circumstances where the data genuinely took a functional form that could not be captured by the model. One circumstance where this would be the case is if there are true discontinuities in learning (Petrov et al., 2005). For instance, in learning experiments that take place over many days, participants may not actually begin each day with the exact same level of performance that they finished with on the previous day (i.e., they may need to readjust to performing the task on each day). This would result in a learning function that is effectively “scalped,” which is a functional form that the current model cannot capture (it would instead tend to smooth over these discontinuities—although the model could certainly be extended in many different ways to account for such data; see, for example, Levi et al., 1997; Yu et al., 2004). Another circumstance that would violate the assumptions of the current approach is if performance improves for a period of time and then proceeds to become worse. This could be the case, for instance, in a long, single-day training experiment where thresholds increase toward the end of training due to fatigue or if participants experience periods of “mind wandering” (McVay & Kane, 2009, 2012); note though that although the nonparametric approach could capture such an occurrence on a block level, it would nonetheless remain difficult to determine how to quantify learning. Finally, the parametric model of perceptual learning is inappropriate for fitting data that is collected via staircase techniques. This is because data collected via staircases (a) rarely includes sufficient spread in stimulus strength to estimate a full psychometric function and (b) what spread in stimulus strength is present is usually highly biased in time, with trials of higher stimulus strength occurring early in the block and trials of lower stimulus strength occurring mainly later in the block (although we note that there are several interesting approaches to examining trial-by-trial data that is generated by staircase techniques, e.g., Ghose, Yang, & Maunsell, 2002; Yang & Maunsell, 2004). There are absolutely a wide range of situations that call for staircase methodology (e.g., when a threshold needs to be estimated very quickly, or when the range of stimulus intensities to present is unknown or could potentially vary substantially between individuals). However, we would suggest that in most learning experiments (which typically involve

thousands of trials and where typical performance is usually reasonably well known) staircases could easily be replaced by methods that are more amenable to the analysis techniques described here. These methods could be augmented by using the technique to adapt level of difficulty so as to maximize learning.

Indeed, it is certainly the case that, given the analysis methods developed and demonstrated here, the methodology that was used to produce the current data can be significantly improved going forward. Specifically, stimulus strength in all cases was drawn from a static uniform range throughout training. This then necessarily meant that as learning proceeded, more and more of the range resulted in ceiling level performance, which is neither ideal with respect to measuring ability (as the far ceiling part of the curve does not help constrain the psychometric function) nor with respect to producing learning (although there is certainly virtue to having some easy trials, it is generally accepted that learning is most efficient when the task is challenging, but doable—when errors are being made, but these errors are informative; Chu, Doshier, & Lu, 2010). It is thus possible that there could be virtue to using the current completely offline analysis approach in an online manner to control the range over which stimulus strength is sampled. This would, in some ways, be the best of both worlds—the uniform random sampling of stimulus strengths allows for continuous estimates of learning—and drifting that range provides a natural way to keep difficulty level constant throughout training (without having to rely on staircases).

Conclusions

Here, we have presented a method to parametrically fit the changes in the psychometric function occurring in perceptual learning studies. Future work with larger datasets will be necessary to identify the best parameterizations of both learning functions and the psychometric functions (and the correct parameterizations may differ across learning domains). Indeed, we note that we ourselves have used a slightly different parameterization in our previous empirical work (Green et al., 2015; Kattner et al., 2017), which used even fewer free parameters, at the cost of additional flexibility. The best tradeoff between these is thus also to be determined. Furthermore, it is likely the case that this approach will be fruitful for analyzing perceptual learning experiments that use alternative designs—such as target present/target absent designs (Ahissar & Hochstein, 1997); however, the parameterization will similarly need to be altered to account for the difference in design. In all, the current data clearly demonstrates that continuous parametric fitting is a flexible tool that

allows for these comparisons to be made using few free parameters and without assuming within-block stationarity.

Keywords: psychometric function, perceptual learning, parametric model

Acknowledgment

This work was supported by the Office of Naval Research grant ONR - N000141712049.

Commercial relationships: none.

Corresponding author: C. Shawn Green.

Email: cshawn.green@wisc.edu.

Address: Department of Psychology, University of Wisconsin-Madison, Madison, WI, USA.

References

- Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, 387(6631), 401–406. <http://doi.org/10.1038/387401a0>
- Astle, A. T., Blighe, A. J., Webb, B. S., & McGraw, P. V. (2015). The effect of normal aging and age-related macular degeneration on perceptual learning. *Journal of Vision*, 15(10):16, 1–16, doi:10.1167/15.10.16. [PubMed] [Article]
- Badiru, A. B. (1992). Computational survey of univariate and multivariate learning curve models. *IEEE Transactions on Engineering Management*, 39(2), 176–188. <http://doi.org/10.1109/17.141275>
- Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research*, 27(6), 953–965. [http://doi.org/10.1016/0042-6989\(87\)90011-3](http://doi.org/10.1016/0042-6989(87)90011-3)
- Beard, B. L., Levi, D. M., & Reich, L. N. (1995). Perceptual learning in parafoveal vision. *Vision Research*, 35(12), 1679–1690. [http://doi.org/10.1016/0042-6989\(94\)00267-P](http://doi.org/10.1016/0042-6989(94)00267-P)
- Bejjanki, V. R., Beck, J. M., Lu, Z.-L. L., & Pouget, A. (2011). Perceptual learning as improved probabilistic inference in early sensory areas. *Nature Neuroscience*, 14(5), 642–648. <http://doi.org/10.1038/nn.2796>
- Casey, M. C., & Sowden, P. T. (2012). Modeling learned categorical perception in human vision. *Neural Networks*, 33, 114–126. <http://doi.org/10.1016/j.neunet.2012.05.001>
- Chu, W., Doshier, B., & Lu, Z.-L. (2010). The rate of perceptual learning at a fixed accuracy threshold is improved by feedback and by mixture with easier trials. *Journal of Vision*, 9(8): 882, doi:10.1167/9.8.882. [Abstract]
- Chung, S. T. L. (2011). Improving reading speed for people with central vision loss through perceptual learning. *Investigative Ophthalmology & Visual Science*, 52(2), 1164–1170. [PubMed] [Article]
- Coates, D. R., & Chung, S. T. L. (2014). Changes across the psychometric function following perceptual learning of an RSVP reading task. *Frontiers in Psychology*, 5(DEC). <http://doi.org/10.3389/fpsyg.2014.01434>
- Crist, R. E., Kapadia, M., Westheimer, G., & Gilbert, C. D. (1997). Perceptual learning of spatial localization: Specificity for orientation, position and context. *Journal of Neurophysiology*, 78(6), 2889–2894.
- Doshier, 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, USA*, 95(23), 13988–13993. <http://doi.org/10.1073/pnas.95.23.13988>
- Doshier, 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. <http://doi.org/10.1111/j.1467-9280.2007.01934.x>
- Doshier, B. A., & Lu, Z.-L. L. (2000). Mechanisms of perceptual attention in precuing of location. *Vision Research*, 40(10–12), 1269–1292. [http://doi.org/10.1016/S0042-6989\(00\)00019-5](http://doi.org/10.1016/S0042-6989(00)00019-5)
- 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. [http://doi.org/10.1016/0042-6989\(93\)90094-D](http://doi.org/10.1016/0042-6989(93)90094-D)
- Fahle, M., & Morgan, M. (1996). No transfer of perceptual learning between similar stimuli in the same retinal position. *Current Biology*, 6(3), 292–297. [http://doi.org/10.1016/S0960-9822\(02\)00479-7](http://doi.org/10.1016/S0960-9822(02)00479-7)
- Fendick, M., & Westheimer, G. (1983). Effects of practice and the separation of test targets on foveal and peripheral stereoacuity. *Vision Research*, 23(2), 145–150. [http://doi.org/10.1016/0042-6989\(83\)90137-2](http://doi.org/10.1016/0042-6989(83)90137-2)
- Fründ, I., Haenel, N. V., & Wichmann, F. A. (2011). Inference for psychometric functions in the presence of nonstationary behavior. *Journal of Vision*, 11(6):16, 1–19, doi:10.1167/11.6.16. [PubMed] [Article]
- Gantz, L., Patel, S. S., Chung, S. T. L., & Harwerth, R.

- S. (2007). Mechanisms of perceptual learning of depth discrimination in random dot stereograms. *Vision Research*, 47(16), 2170–2178. <http://doi.org/10.1016/j.visres.2007.04.014>
- Ghose, G. M., Yang, T., & Maunsell, J. H. R. (2002). Physiological correlates of perceptual learning in monkey V1 and V2. *Journal of Neurophysiology*, 87(4), 1867–1888. <http://doi.org/10.1152/jn.00690.2001>
- Green, C. S., Kattner, F., Siegel, M. H., Kersten, D., & Schrater, P. R. (2015). Differences in perceptual learning transfer as a function of training task. *Journal of Vision*, 15(10):5, 1–14, doi:10.1167/15.10.5. [PubMed] [Article]
- Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96(3), 280–301. <http://doi.org/10.1016/j.bandl.2005.06.001>
- Heathcote, A., Brown, S., & Mewhort, D. J. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, 7(2), 185–207. <http://doi.org/10.3758/BF03212979>
- Herzog, M. H., & Fahle, M. (1997). The role of feedback in learning a vernier discrimination task. *Vision Research*, 37(15), 2133–2141.
- Herzog, M. H., & Fahle, M. (1998). Modeling perceptual learning: Difficulties and how they can be overcome. *Biological Cybernetics*, 78(2), 107–117. <http://doi.org/10.1007/s004220050418>
- Herzog, M. H., & Fahle, M. (1999). Effects of biased feedback on learning and deciding in a vernier discrimination task. *Vision Research*, 39, 4232–4243.
- Jacobs, R. A., & Kruschke, J. K. (2011). Bayesian learning theory applied to human cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(1), 8–21. <http://doi.org/10.1002/wcs.80>
- Jones, P. R., Moore, D. R., & Amitay, S. (2015). The role of response bias in perceptual learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2015(5), 1456–1470. <http://doi.org/10.1037/xlm0000111>
- 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. <http://doi.org/10.1016/j.cub.2017.01.046>
- Klein, S. A. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics*, 63(8), 1421–1455. <http://doi.org/10.3758/BF03194552>
- Law, C.-T., & Gold, J. I. (2009). Reinforcement learning can account for associative and perceptual learning on a visual-decision task. *Nature Neuroscience*, 12(5), 655–663. <http://doi.org/10.1038/nn.2304>
- Levi, D. M., Polat, U., & Hu, Y. S. (1997). Improvement in Vernier acuity in adults with amblyopia: Practice makes better. *Investigative Ophthalmology and Visual Science*, 38(8), 1493–1510. [PubMed] [Article]
- Liu, J., Doshier, B. A., & Lu, Z.-L. (2014). Modeling trial by trial and block feedback in perceptual learning. *Vision Research*, 99, 46–56. <http://doi.org/10.1016/j.visres.2014.01.001>
- Liu, Z., & Weinshall, D. (2000). Mechanisms of generalization in perceptual learning. *Vision Research*, 40(1), 97–109. [http://doi.org/10.1016/S0042-6989\(99\)00140-6](http://doi.org/10.1016/S0042-6989(99)00140-6)
- Lu, Z.-L., Hua, T., Huang, C.-B., Zhou, Y., & Doshier, B. A. (2011). Visual perceptual learning. *Neurobiology of Learning and Memory*, 95(2), 145–151. <http://doi.org/10.1016/j.nlm.2010.09.010>
- Matthews, N., Liu, Z., Geesaman, B. J., & Qian, N. (1999). Perceptual learning on orientation and direction discrimination. *Vision Research*, 39(22), 3692–3701.
- Mazur, J. E., & Hastie, R. (1978). Learning as accumulation: A reexamination of the learning curve. *Psychological Bulletin*, 85(6), 1256–1274. <http://doi.org/10.1037/0033-2909.85.6.1256>
- McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(1), 196–204. <http://doi.org/10.1037/a0014104>
- McVay, J. C., & Kane, M. J. (2012). Drifting from slow to “d’oh!”: Working memory capacity and mind wandering predict extreme reaction times and executive control errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(3), 525–549. <http://doi.org/10.1037/a0025896>
- Michel, M. M., & Jacobs, R. A. (2007). Parameter learning but not structure learning: A Bayesian network model of constraints on early perceptual learning. *Journal of Vision*, 7(1):4, 1–18, doi:10.1167/7.1.4. [PubMed] [Article]
- Moscattelli, A., Mezzetti, M., & Lacquaniti, F. (2012). Modeling psychophysical data at the population-level: The generalized linear mixed model. *Journal of Vision*, 12(11):26, 1–17, doi:10.1167/12.11.26. [PubMed] [Article]
- Newell, A., & Rosenbloom, P. S. (1981). Mechanisms

- of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–51). Hillsdale, NJ: Lawrence Erlbaum.
- Petrov, A. A., Doshier, B. A., & Lu, Z.-L. (2005). The dynamics of perceptual learning: An incremental reweighting model. *Psychological Review*, *112*(4), 715–743. <http://doi.org/10.1037/0033-295X.112.4.715>
- Petrov, A. A., Doshier, B. A., & Lu, Z. L. (2006). Perceptual learning without feedback in non-stationary contexts: Data and model. *Vision Research*, *46*(19), 3177–3197. <http://doi.org/10.1016/j.visres.2006.03.022>
- Poggio, T., Fahle, M., & Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. *Science*, *256*(5059), 1018–1021. <http://doi.org/10.1126/science.1589770>
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–408. <http://doi.org/10.1037/h0042519>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. <http://doi.org/10.1038/323533a0>
- Seitz, A. R., Nanez, J. E., Holloway, S., Tsushima, Y., & Watanabe, T. (2006). Two cases requiring external reinforcement in perceptual learning. *Journal of Vision*, *6*(9):9, 966–973. [PubMed] [Article]
- Snell, N., Kattner, F., Rokers, B., & Green, C. S. (2015). Orientation transfer in vernier and stereoacuity training. *PLoS One*, *10*(12), e0145770. <http://doi.org/10.1371/journal.pone.0145770>
- Sotiropoulos, G., Seitz, A. R., & Seris, P. (2011). Changing expectations about speed alters perceived motion direction. *Current Biology*, *21*(21), R883–R884. <http://doi.org/10.1016/j.cub.2011.09.013>
- Spratling, M. W., & Johnson, M. H. (2001). Dendritic inhibition enhances neural coding properties. *Cerebral Cortex*, *11*(12), 1144–1149. <http://doi.org/10.1093/cercor/11.12.1144>
- Spratling, M. W., & Johnson, M. H. (2006). A feedback model of perceptual learning and categorization. *Visual Cognition*, *13*(2), 129–165. <http://doi.org/10.1080/13506280500168562>
- Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research*, *35*(17), 2503–2522. [http://doi.org/10.1016/0042-6989\(95\)00016-X](http://doi.org/10.1016/0042-6989(95)00016-X)
- Vaina, L. M., Sundaeswaran, V., & Harris, J. G. (1995). Learning to ignore: Psychophysics and computational modeling of fast learning of direction in noisy motion stimuli. *Cognitive Brain Research*, *2*(3), 155–163. [http://doi.org/10.1016/0926-6410\(95\)90004-7](http://doi.org/10.1016/0926-6410(95)90004-7)
- Yang, T., & Maunsell, J. H. R. (2004). The effect of perceptual learning on neuronal responses in monkey visual area V4. *The Journal of Neuroscience*, *24*(7), 1617–1626. <http://doi.org/10.1523/JNEUROSCI.4442-03.2004>
- Yu, C., Klein, S. A., & Levi, D. M. (2004). Perceptual learning in contrast discrimination and the (minimal) role of context. *Journal of Vision*, *4*(3):4, 169–182, doi:10.1167/4.3.4. [PubMed] [Article]
- Zhaoping, L., Herzog, M. H., & Dayan, P. (2003). Nonlinear ideal observation and recurrent preprocessing in perceptual learning. *Network (Bristol, England)*, *14*(2), 233–247. <http://doi.org/10.1088/0954-898X/14/2/304>