Journal of Cognitive Enhancement
https://doi.org/10.1007/s41465-024-00303-3

REVIEW

Modeling Perceptual Learning as a Continuous Function of Time-on-Task Increases Theoretical Specificity and Statistical Power

Aaron Cochrane[1] · Zhong-Lin Lu[2,3] · C. Shawn Green[4]

Received: 15 November 2023 / Accepted: 2 July 2024

Abstract

Drawing inferences from data relies on satisfying the assumptions of analytical methods. Yet, the study of many psychological processes that involve change over time (e.g., learning) instead uses methods that assume a lack of change over time (e.g., within- "block" averages). Recent research has demonstrated a variety of theoretical and empirical benefits in aligning the assumptions about the generative process (e.g., changing perceptual sensitivity due to learning) with assumptions in analyses (e.g., changing estimate of sensitivity as a continuous function of time). In this review, we examine methods for estimating performance on a trial-to-trial basis as it changes due to learning. We then explore applications of these methods, including increased efficiency and statistical power, as well as the ability to more effectively investigate questions of learning generalization and individual differences. We highlight the applicability and utility of continuous-time models in perceptual learning, cognitive training, and beyond.

Keywords Perceptual learning · Transfer · Learning to learn · Parametric models · Non-parametric models

Introduction

Learning is the improvement in skills, abilities, or knowledge as a function of experience. Critically, most models of learning consider the relationship between experience and learning to be, at a minimum, monotonic. In other words, most theoretical models instantiate the idea that each piece of experience should produce either no effect or some concomitant degree of improvement in performance.

Given the prevalence of an experience-dependent view of learning, it is perhaps surprising how commonly learning is, in practice, studied via approaches that implicitly assume that the process linking experience with changes in performance has major periods of stationarity (i.e., where performance is not changing with experience). Consider, for example, a perceptual learning task wherein on each of 1000 trials participants are shown two short intervals of random dot motion. The participants' task is to indicate whether the net direction of the dots in the two intervals was the same or was different by 3 degrees. A typical way of analyzing this data would be to first divide the 1000 trials into smaller chunks or blocks (e.g.,100-trial blocks; noting these blocks frequently align with "days" or "sessions" in multi-session training experiments). A single metric of performance (e.g., d') would then be calculated for each block of 100 trials. Finally, those block-aggregated metrics would be utilized in all subsequent analyses (e.g., in calculating total learning, in fitting a function to the block metrics, in assessing transfer index).

Other approaches from the same general family might involve, for instance, using an adaptive staircase procedure over each of the 100-trial blocks to find the minimum motion direction difference that is required for participants to reach a certain threshold of performance in that block. Or perhaps rather than an adaptive algorithm, a method of constant stimuli is utilized in each 100-trial block (e.g., five values of motion direction differences that are each repeated 20 times per block). Like in the top example, in both cases above, the procedure will, in the end, result in a single metric of performance for each 100-trial block.

How, then, are these approaches instantiating the idea that the link between experience and changes in performance

Aaron Cochrane
aaron_cochrane@brown.edu

1 Brown University, Providence, RI, USA

2 NYU Shanghai, Shanghai, China

3 New York University, New York, NY, USA

4 University of Wisconsin – Madison, Madison, WI, USA

Published online: 19 July 2024

have periods of stationarity? Implicitly, by aggregating over trials in blocks, a person is making one of two assumptions. One possible assumption is that the trials within a block are independent and identically distributed (*iid*). In essence, one is assuming that performance is stable and ergodic within that block of trials (Molenaar, 2008). If the first trial in a block is theoretically substitutable for the 100th trial of the block, taking the average will simply reduce the influence of any extraneous noise that could be corrupting performance. A second possible assumption is that the trials within a block are not *iid*, but any deviations from this assumption do not matter for the questions at hand.

Here, we present evidence that neither of the assumptions above is typically borne out in empirical data. We thus argue for the need to move away from aggregation-based analyses of learning and move toward more continuous-time approaches. We start by providing an overview of continuous-time approaches to examining behavioral data. We next provide examples establishing that across a wide range of tasks (stimulus domains, task lengths), performance is not stationary; it is instead changing through time in a continuous and systematic fashion that can be captured. We then turn to the weightier question of "Why should researchers care if the assumptions underlying averaging are violated?" In this we review, a variety of recent results demonstrating how inappropriately aggregating over time would have resulted in inferences that range from incomplete to misleading. We then briefly discuss additional benefits of utilizing continuous-time approaches (e.g., with respect to power, with respect to producing results that are less dependent on methodological choices, and with respect to identification of mechanisms). Finally, we conclude by considering future directions—including how to ease the adoption of continuous-time analysis approaches.

Toward a Definition of Continuous-Time Models

In the majority of cases where researchers have studied learning, aggregation-based analyses, rather than those that model trajectories on the level of individual time points and individual learners, have been the norm (Brown & Heathcote, 2003; Crossman, 1959; A. Newell & Rosenbloom, 1981; Snoddy, 1926). The literature, though, is sprinkled with some examples of researchers examining performance on a completely granular, trial-by-trial level. For instance, in the quest to determine whether learning occurred gradually (i.e., "Law of Effect") or all-at-once (i.e., via "insight"), early work by Thorndike frequently involved plotting and examining data at the trial by trial level (Woodworth & Thorndike, 1901; see also Dayan & Daw, 2008). Similarly, in Harlow (1949), differences in individual trials' performance were used to demonstrate accelerating learning and the existence of learning sets (i.e., generalization due to learning structure of task demands independently of specific details). And more recently, Newell et al. (2009) argued for a finer-grained analysis of learning which may reveal trajectories differing greatly from conventional functions (e.g., power; A. Newell & Rosenbloom, 1981). In this vein, trial-by-trial performance in learning (Gallistel et al., 2004) has been used to argue for a step-function-like pattern of change.

However, while previous arguments for a consideration of learning at the level of individual trials and learners have largely been limited to contexts in which researchers believed the behaviors under study might improve very rapidly (e.g., with "insight") or for behavioral paradigms for which reinforcement learning algorithms (e.g., Dayan & Daw, 2008; Montague & Sejnowski, 1994) are tractable, we believe the implications should be applied far beyond these limits. Just as the use of Bayesian and mixed-effects statistical models has seen a rapid growth in the last several decades due to increasing computational resources and software developments, we propose that by-trial and by-participant models can be developed and should be increasingly applied to studying perceptual and cognitive learning. Such approaches can take several forms, of which only a restricted set is the focus of this manuscript.

Models Parameterizing Change in Performance

Here, we primarily consider a specific case of by-trial and by-participant models, specifically, parametric models of change. In this context, the index of performance is the predicted value of some saturating function, such as a power or exponential function. In the "Discussion," we will also briefly explore approaches that do not make such assumptions. One intuitive and straightforward method for fitting parametric models to by-trial and by-participant data, for example, a measure like absolute estimation error, is to use an ordinary least-squares regression model that predicts the error as a function of trial number for each participant. It is worth noting that linear functions of trial number are generally not theoretically supported and are likely to produce biased and/or impossible estimates in practice (see Cochrane & Green, 2021b and non-parametric section below). Consequently, saturating functions are necessary (Dosher & Lu, 2007; K. M. Newell et al., 2009). As an example, we use an exponential function with parameters start, rate, and asymptote:

$$\text{absolute error} \sim \text{start} + (\text{asymptote} - \text{start})^{(\text{rate} * \text{timepoint})}$$

While this may be a reasonable approach under certain circumstances, it is important to acknowledge that meaningful research is rarely so simple. In many cases, our interest goes beyond predicting the overall mean. We might instead be interested in predicting, for example, the psychometric function threshold using the exponential function for each trial (see Fig. 1). We may also be interested in the effect of various covariates on the learning process (e.g., experimental condition or concurrently-measured electrophysiological responses). Clearly, the simple regression case is unduly restrictive in these scenarios. Two extensions are warranted. Firstly, the nonlinear regression aims to predict the threshold for each trial rather than the overall mean:

$$\text{threshold} \sim \text{start} + (\text{asymptote} - \text{start})^{(\text{rate}*\text{timepoint})}$$

In the second extension, the parameters of change can be further influenced by other factors within a standard linear regression design matrix. This includes main effects and interactions of categorical or numeric predictors:

$$\text{start} \sim \text{design\_matrix}$$

$$\text{rate} \sim \text{design\_matrix}$$

$$\text{asymptote} \sim \text{design\_matrix}$$

Luckily, each of these extensions is quite simple to implement in modern statistical software (Bürkner, 2017; Cochrane, 2020). A generative model, with the threshold as a function of time, its covariates, some other psychometric function parameter(s) (e.g., lapse rate; Weibull shape; logistic bias), and some stimulus value (e.g., stimulus strength), is constructed to generate the performance predictions (e.g., 80% probability of a "clockwise" response; see Fig. 1) based on a set of parameters. A likelihood function is then used to link the prediction of the model to the trial-by-trial performance in the experiment.

A key problem with such models, however, is their complexity relative to the quantity of data. Fortunately, compared to conventional methods that use maximum-likelihood estimation for blocks of trials (and must implement compromises regarding the balance between the robustness of estimates and their temporal precision), Bayesian mixed-effects methods allow for the use of priors and random-effects structures to provide stability and robustness to model fits (Cochrane et al., 2022).

The parametric model can also serve as the basis for adaptive measurement techniques. With a parametric model of how perceptual sensitivity systematically changes with the observer's state over time, Zhao et al. (2019) developed the quick change-detection (qCD) method, which accurately, precisely, and efficiently assesses perceptual sensitivity in every trial. Using the Bayesian active learning framework (Lesmes, 2010; Lesmes et al., 2006, 2015; Watson & Pelli, 1983), qCD selects the optimal stimulus that would yield the most new information about perceptual sensitivity in the next trial and updates a joint probability distribution of the parameters of a model of perceptual sensitivity change over time.

The Bayesian Inference Procedure (BIP) that is used to make inference about the learning processes of each individual subject independently can be further improved with hierarchical Bayesian modeling (HBM). HBM is a generative model framework that utilizes Bayes' rule to quantify the joint posterior distribution of test-, subject-, and population-level parameters and hyperparameters (Kruschke, 2011; Lee, 2011; Rouder & Lu, 2005; Wilson

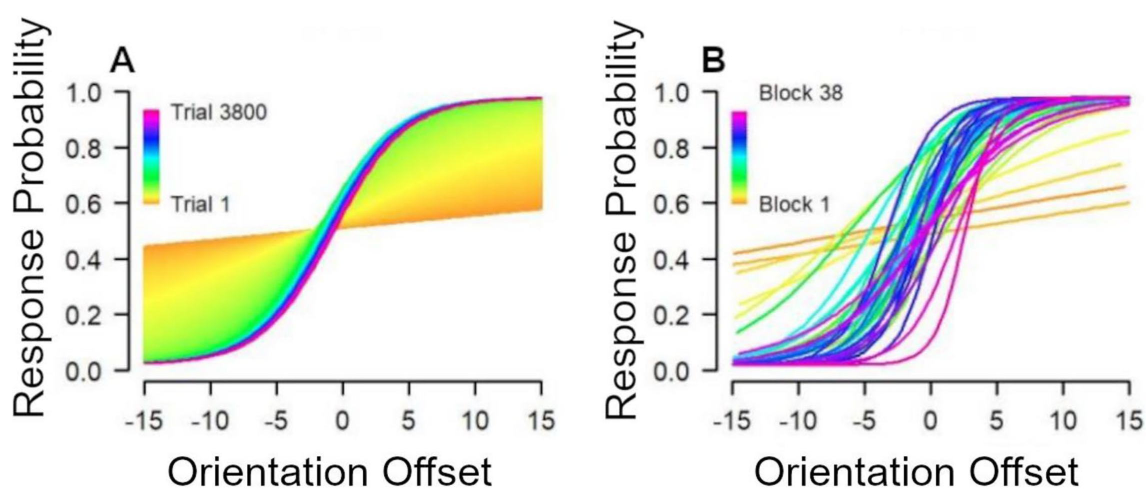

Fig. 1 Comparing continuous-time and block-based approaches to fitting psychophysical performance. **A** A logistic psychometric function, which links a stimulus strength (here offset from a reference orientation) to probability of a binary response, may be estimated as a continuous saturating function of time, here moving from reasonably poor performance on trial one (orange color) to much better performance on trial 3800 (purple color). **B** This differs from the standard approach of fitting a separate psychometric function to separate blocks of trials; here, 38 independent logistic functions are fit to 100-trial blocks. Figure adapted from Kattner et al., (2017a, 2017b)

et al., 2020). Because experimental designs are almost always hierarchical (e.g., trials nested within sessions nested within subjects nested within a population), statistical methods accounting for this hierarchy are theoretically informed and practically useful. Rather than assuming across-subject homogeneity (i.e., aggregating across subjects; A. Newell & Rosenbloom, 1981) or assuming independence (i.e., estimating subjects' performance separately from one another), leveraging the hierarchical structure allows for a systematic and powerful combination of information across levels. Estimated models then provide joint posterior distribution of parameters at each of the nested levels (Kruschke & Vanpaemel, 2015; Zhao et al., 2021). Such a procedure reduces posterior variances (i.e., uncertainty) both through efficient parcellation of variance (Song et al., 2020) and increased robustness to smaller amounts of data in certain levels ("shrinkage;" Rouder & Lu, 2005; Rouder et al., 2003). Concretely, individual-level estimates are less likely to be outliers due to noisy data; random-effects structures push estimated coefficients into closer conformity with the population level distribution, thereby providing robustness in the face of noisy data. Such benefits are compounded when using Bayesian methods. By facilitating the seamless integration of prior expectations, Bayesian analyses can reduce the chance of estimated values being driven by noise (e.g., a researcher may expect "orientation discrimination threshold is often around 5 degrees using this paradigm, and it cannot be smaller than zero or larger than ninety"). In conjunction with unified mixed-effects methods, Bayesian methods can increase the power to detect effects of interest while mitigating the influence of noise or overfitting.

Practically speaking, when constructing a linear mixed-effects model structure to predict asymptotic performance (as in *asymptote ~ design_matrix* above), a similar approach must be taken as in conventional mixed-effects models. This involves identifying parameters of interest to estimate (e.g., contrasts between experimental conditions) or between-participant covariates (e.g., age) and including these in the fixed-effects structure. The random-effects structure can then be defined by including by-group effects (e.g., by-site or by-participant intercepts) as well as coefficients modifying those effects (e.g., by-participant slope for stimulus type). For instance, in a recent paper (Cochrane et al., 2023a, c) examining changes in perceptual sensitivity as measured by an evidence-accumulation model's drift rate, asymptotic drift rate was modeled as a function of fixed effects (an overall intercept and a coefficient associated with a covariate, whether a participant was an action video game player or not) as well as random effects (by-participant intercepts, and by-participant coefficients for the function of stimulus strength on drift rate). If the effect of stimulus strength on drift rate was desired, then it could have been included in the fixed-effects structure, but in this context, it was simply being controlled for.

Comparison of Continuous-Time Models to Conventional Approaches

Unlike continuous-time analyses of learning, conventional aggregation-based methods typically rely on an index of performance extracted from a set of trials. For example, performance sensitivity (e.g., d') in a detection task may be calculated for each block of 100 trials. Other common performance indices include estimated or staircase-converged thresholds (Gold & Ding, 2013; Watson & Pelli, 1983), percent correct, and, more recently, evidence-accumulation model indices have been increasingly utilized (Cochrane et al., 2023c; Eckhoff et al., 2008).

A fundamental assumption underlies all such aggregation-based models that the process(es) generating the data at the beginning of a block is identical to the process(es) generating the data at the end of the block. We argue, however, that in studies concerning learning, the opposite assumption is usually true. Learning often changes the generative process even within a block of trials (e.g., increasing perceptual sensitivity). When a method produces a single performance value to represent many trials, it implements the testable assumption that the perceptual process of interest remains unchanged from the beginning to the end of that block. We propose that such an assumption should be paradigmatically tested, for instance, via direct model comparisons. When such comparisons have been conducted, they have consistently provide evidence for continuous changes (Cochrane & Green, 2021a; Cochrane et al., 2023b; Dale et al., 2018; Kattner et al., 2017a, 2017b).

Furthermore, estimating parameters for each block independently increases the number of estimated parameters, increasing risk of data overfitting when using conventional methods. We have shown that by fitting trajectories with a small number of parameters, rather than independent psychometric functions, such overfitting can be avoided (e.g., out-of-sample predictiveness is increased; Kattner et al., 2017b).

In another comparison with standard methods, Zhao et al. (2019) demonstrated the efficacy of the qCD method in dark adaptation with computer simulations of an eight-alternative forced-choice (8AFC) task. They showed that one run of the qCD method could estimate the dark adaptation curve with better accuracy and precision than ten runs of the quick forced choice (qFC; Lesmes et al., 2015) and staircase methods. Moreover, the dark adaptation curve obtained from one qCD run in a psychophysics experiment closely matched the average of four qFC runs (RMSE = 0.076 $\log_{10}$ units). Zhang et al. (2019) implemented and tested the qCD method in assessing the learning curve in a four-alternative forced-choice global motion direction identification task, both in simulations and a psychophysical experiment. Simulations revealed that the accuracy and precision (standard deviation or confidence bounds) of the estimated learning curves from the qCD method outperformed those obtained by the staircase and random stimulus selection methods.

By using continuous estimation, performance estimates become more efficient and less sensitive to noise. In comparison to more-standard methods, there is a reduced likelihood of fitting noise (Kattner et al., 2017a, 2017b), thus allowing for performance estimation with higher confidence (Zhang et al., 2019) and with fewer trials (Zhao et al., 2019). However, it is important to note that such conclusions come with caveats. Employing a parametric approach to estimation assumes the appropriateness of the parameterization (see "Functional Forms of Learning" and "Estimating Nonparametric Changes" below). If variations that deviate from a function are indeed noise, such as transient reductions in performance due to attentional lapses, then a parametric function provides robustness against such noise. On the other hand, if these variations are of interest, as in studies addressing attentional lapses, then a parametric function without explicit modeling of attentional lapse may hinder the detection of relevant phenomena. This emphasizes the trade-off between temporal precision and robustness to noise, which is also inherent in block-based methods.

In conclusion, the available empirical data aligns with the theoretical perspective highlighted at the beginning of the introduction—learning in most cases proceeds continuously with experience, rather than being characterized by long periods of stationarity that happen to coincide with reasonably arbitrary blocks of trials (i.e., stationary between trial 1 and 100, but then changes between trial 100 and trial 101, only to be stationary again between trial 101 and 200).

Applications and Implications

Both in principle and in practice, the available empirical data convincingly demonstrates that within-block stationarity of processes should not be assumed (it is worth noting that model comparisons can indicate evidence for a stationary generative process if it, in fact, exists). The question then arises: what is the "gain" associated with modelling data in a continuous-time fashion and what is the "cost" of not doing such modelling? Given the better alignment with both theory and real-world behavior, our *a priori* expectation is that implementing continuous-time models would provide benefits for drawing inferences from data. Indeed, as we will review below, the applications of continuous-time models are widespread, and the implications are straightforward: the goal is to align the assumed generative process (i.e., change in a process due to experience) with the interpretative method.

Here, we outline several examples of how continuous-time inferences provide insights that non-continuous-time estimates would not allow.

Functional Forms of Learning, and Their Implications for Underlying Processes of Change

One fundamental question in the study of learning concerns the underlying function governing change. Does learning progress in plateaus or are there moments of "insight?" Does learning accelerate over time, decelerate, or maintain a constant proportional rate? The answers to these questions are not only of intrinsic interest but also serve as critical tests of theory—as various theoretical accounts of learning produce characteristic functional forms. Considering the functional forms of change can provide mechanistic insights as well as practical suggestions. Historically, such comparisons led to the so-called "power law of learning," where change was thought to follow a saturating 3-parameter power function of time (Crossman, 1959; Logan, 1988; A. Newell & Rosenbloom, 1981).

However, the "power law of learning" has been challenged, due to the typical aggregation of data across learners, blocks of trials, or both. It has been argued that "power law of learning" has been "overturned" in favor of alternative functional forms (Brown & Heathcote, 2003; Dosher & Lu, 2007; Gallistel et al., 2004; Heathcote et al., 2000). For instance, multiple studies have shown that when fitting at the individual (rather than group aggregate), the best-fitting functional form is a 3-parameter exponential function of change rather than power function (see Equation above; Dosher & Lu, 2007; Heathcote et al., 2000). Exponential functions represent constant proportions of change for each unit of time, while power functions lead to a decreasing proportion of change, indicating deceleration and an interaction of multiple forms of change.

Yet, while these critiques have shown that averaging exponential forms across learners results in a group-averaged trajectory best fit by a power function, less work has been reported on the effects of averaging across blocks of trials. However, recent studies, such as our work using by-participant and by-trial models, have shown that functions from the exponential family fit better than power functions in two perceptual learning paradigms (dot-motion direction discrimination and texture oddball detection; Cochrane & Green, 2021a). Furthermore, the temporal precision allowed by trial-to-trial estimates revealed that motion perceptual learning involved an initially increasing rate of learning rather than a decreasing rate (resembling a sigmoid-like function; see also Gallistel et al., 2004).

Aggregating performance data over many trials, such as with staircase runs or with psychometric function estimates, reduces the temporal precision of performance estimates, which can obscure important features of learning. For instance, understanding the processes of change requires considering both within-block and between-block learning effects (Yang et al., 2022); interactions may occur between forgetting and consolidation (between-block effects) as well as between trial-to-trial learning and fatigue (within-block effects; see Fig. 2). The examination of within-block learning can reveal whether it's part of a single global trajectory or an isolated member of a sequence of within-day trajectories (Cochrane et al., 2023c; Yang et al., 2022).

In Yang et al. (2022), a dataset of 49 adult subjects in seven perceptual learning tasks, each trained in five daily sessions, was analyzed. The authors fit a series of models to the average block-by-block learning curves. The full model contained 24 parameters, including those for general learning, offline gain/forgetting, relearning, adaptation, and rates of change within sessions. The best-fitting models for various tasks included different combinations of these parameters, highlighting the task-specific nature of learning processes.

Identifying within-day and between-day continuities or disjunctions can provide process-level inferences about the timescales of change and their dependence on between-session breaks. Cochrane et al., (2023c) used trial-to-trial trajectories of change in drift diffusion model parameters over 4 days of perceptual learning data, comparing models allowing for various forms of between-day continuities and disjunctions. Perceptual sensitivity (i.e., drift rate) changed as a relatively simple saturating function of overall experience, independent of between-day breaks. In contrast, response boundary, associated with caution or speed-accuracy trade-off, demonstrated between-day changes with varying patterns of between-day and within-day changes across days and across learners.

Mechanisms of Generalization that May Not Be Dissociable When Assuming Stationarity

Beyond advancing our understanding of the learning processes themselves, the increased temporal precision of performance estimates offered by continuous-time approaches can play a crucial role in detecting and interpreting the effects of specificity and generalization in perceptual learning. In most learning domains, a core question revolves around the extent to which learning in one task generalizes to new, untrained tasks. Perceptual learning, especially when utilizing single low-variability training tasks, often demonstrates high specificity (e.g., initial training does not benefit performance in a changed task; Ahissar & Hochstein, 1997; Ball & Sekuler, 1987; Fahle, 2005; Fiorentini & Berardi, 1980; Jeter et al., 2009). However, proposed applications of perceptual learning rely on the promise of generalization from initial training to a wider variety of tasks and performance demands, and such generalization has been observed under certain conditions (Deveau et al., 2013; Jeter et al., 2010; Kattner et al., 2017a; Xiao et al., 2008).

Yet, while the field of perceptual learning has greatly expanded its knowledge of how to produce additional generalization of learning, block-based analyses make it difficult or impossible to adjudicate between two theoretically distinct forms of generalization. The first is element-level transfer (i.e., due to shared features of training task and tested generalization; Wimer, 1964; Woodworth & Thorndike,

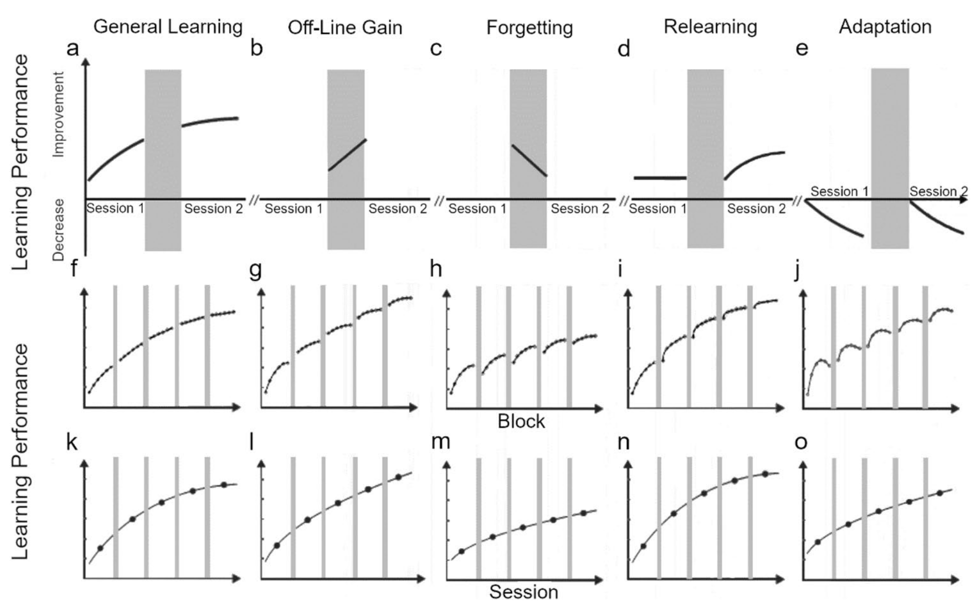

Fig. 2 Performance change may proceed, as a function of time, in a variety of distinct ways. Processes of change may include learning as a function of overall experience (**a**, **f**, **k**), offline gains (**b**, **g**, **l**), forgetting (**c**, **h**, **m**), relearning (**d**, **i**, **n**), or adaptation (**e**, **j**, **o**). While the influence of each of these processes is able to be identified when examining performance on a within-block timescale (**f**, **g**, **h**, **i**, **j**), these processes are not dissociable when examining performance using only estimates of full sessions (**k**, **l**, **m**, **n**, **o**). Figure adapted from Yang et al. (2022)

1901), where participants trained on task A immediately perform a new task B at a high level due to shared features of the training task and the generalization task. The second form is an improved ability to learn new tasks (e.g., due to learning tasks' hierarchical structures or due to domain-general improvements in learning ability; Bejjanki et al., 2011, 2014; Green et al., 2010), where participants trained on task A may not show immediate benefits on task B but learn task B more quickly than if they had not undergone training on task A. These two possibilities of generalization are characterized by different trajectories of learning within a generalization task. Element-level generalization is identified via immediate benefits during the generalization task, while improved ability to learn is characterized by an increased rate of learning during generalization.

Kattner et al., (2017a) demonstrated that these two forms of generalization were conflated when using typical aggregation-based measurements of generalization (see Fig. 3). In contrast, by assessing the time course of generalization using psychometric threshold as a function of time (Cochrane, 2020; Kattner et al., 2017b), the two possible routes to generalization were dissociable. Kattner et al., (2017a) showed that perceptual training on a series of four tasks did not lead to greater immediate transfer than an equivalent amount of training on a single task, but it did lead to faster learning during generalization. In this case, by-trial trajectories of learning during generalization were necessary to identify the form of generalization.

Continuous-time models can also increase precision in more conventional measures of generalization in perceptual learning studies (see chapter 3 of Dosher & Lu, 2020 for a review), such as the transfer index (TI):

$$TI = 1 - (T_{B1} - T_{Bend})/(T_{A1} - T_{Aend})$$

where T_{A1} and T_{Aend} are the estimated thresholds in the beginning and end of training, T_{B1} and T_{Bend} are the estimated thresholds in the beginning and end of the transfer. The accuracy and precision of the initial and final thresholds in the learning phase and the initial threshold in the transfer phase are critical for computing the transfer index (Ahissar & Hochstein, 1997; Dosher & Lu, 2007; Jeter et al., 2010; Liu et al. 2012). Zhang et al. (2019) showed that the continuous qCD method provided highly accurate and precise estimates of initial thresholds and transfer indices, especially in cases of rapid learning, whereas the staircase method was much worse.

The study of perceptual and cognitive training has led to the convention of occasionally separating "training" blocks of trial with feedback from "test" blocks of trials without feedback, often with the assumption that "pre-test" and "post-test" are free of learning artifacts, and learning only happens during "training." However, the strict dichotomy is

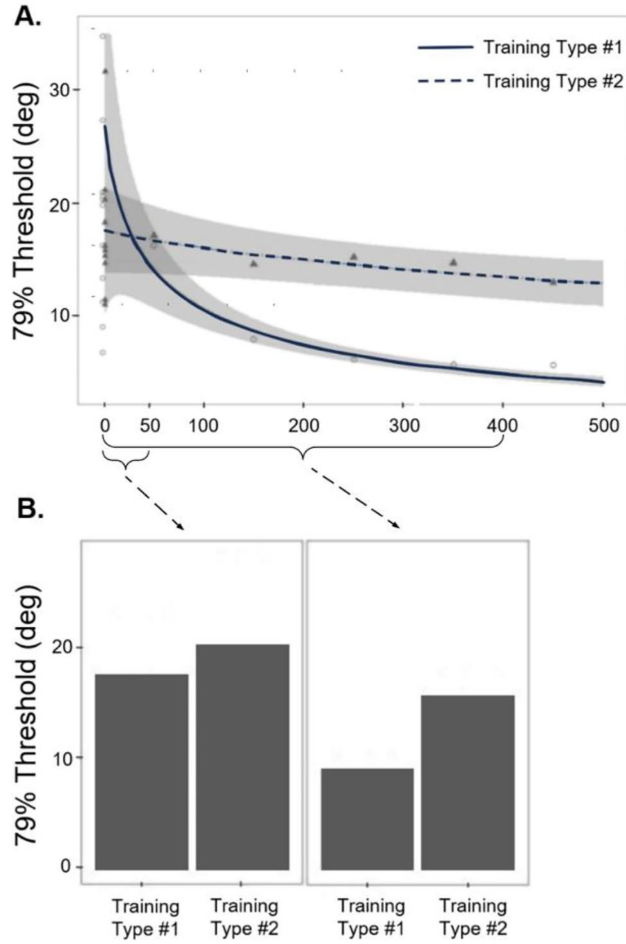

Fig. 3 Generalization of learning may occur due to immediate benefits on a subsequent task or by accelerating learning a subsequent task. While these two routes to generalization are dissociable when using continuous-time analyses (panel **A**), they may not be dissociable when aggregating performance over many trials (panel **B**). In fact, when aggregating performance, choice of block size (which is often not theoretically motivated) may bias inferences regarding generalization (panel **B**; 400-trial blocks show generalization while 50-trial blocks do not show generalization). By estimating continuous parameters of change rather than blocks, it can be seen in panel **A** that performance did not immediately generalize, but instead improved faster after training type 2 than after training type 1. Figure adapted from Kattner et al., (2017a)

challenged by various examples of perceptual learning without feedback (Fahle & Edelman, 1993; Karni & Sagi, 1993; Shiu & Pashler, 1994), and the assumption should be open to question (Seitz et al., 2006, 2009; Watanabe et al., 2001). Furthermore, conventional assumptions have prioritized the interpretation of performance "free of learning artifacts," relegating performance "during learning" to a secondary position. We propose breaking the dichotomy and reevaluating these categorizations of performance.

Lastly, the idea that generalization can manifest as differences in learning in some cases requires shifts in broad methodological approaches. For instance, if a certain type of training is expected to impact the ability to learn new tasks, it may be important to avoid a pre-test altogether, as learning during the pre-test would reduce the extent to which differences in learning at the post-test could be observed. Instead, the key variable should be the learning rate at the post-test, contrasting a treatment with a control. In short, it means a switch from a pre-test → training (treatment arm and control arm) → post-test design, to a training (treatment arm and control arm) → post-test (analyze learning rate) design. This shift in approach was evident in Zhang et al. (2021), where the benefits of training with action video games (on both working memory task performance and perceptual task performance) substantially manifested as an enhanced ability to learn new tasks after action video game training.

Meaningful Individual Differences that Include More Process Specificity

By drawing inferences about behavior using components of performance trajectories rather than overall aggregated values, we can identify dissociable individual differences (Cochrane & Green, 2021a, 2021b; Molenaar, 2004). This approach is particularly valuable in the study of learning, where it helps distinguish between the rates and magnitudes of learning, which might otherwise be conflated in inter-individual analyses (see Fig. 4).

Dale et al. (2021) provided an example by training participants in two different perceptual learning tasks and extracting components of trial-to-trial learning trajectories. They found that the learning rate and asymptotic ability of learners were correlated across visual tasks, providing evidence for a common factor. The values of the common learning rate factor were correlated with a variety of other measures, such as sensitivity to punishment. In contrast, time-dependent aspects of perceptual learning did not show reliable correlations with fluid intelligence scores or conscientiousness, addressing concerns about potential confounds. By examining these correlations on separate components of learning, they were able to test distinct hypotheses about individual differences in perceptual learning ability. For instance, speed of processing was positively (although not always reliably)

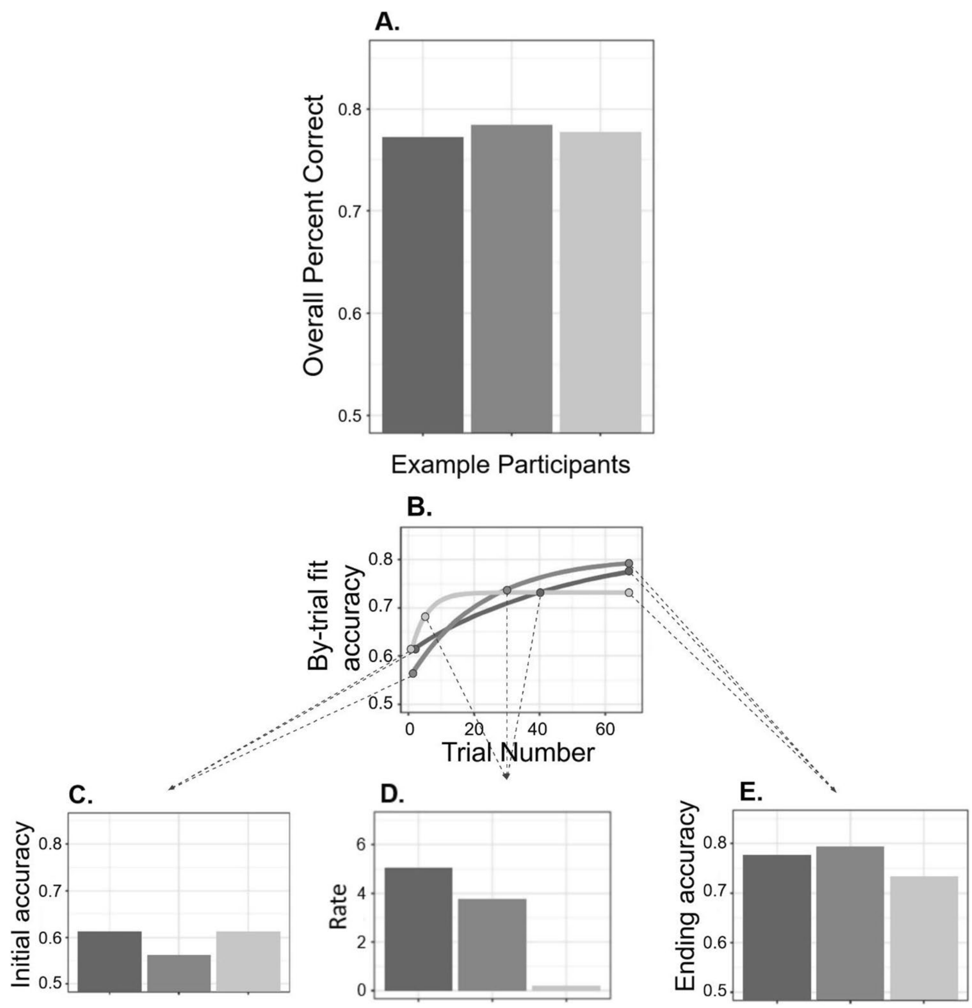

Fig. 4 Impact of continuous-time analyses on consideration of individual differences. When examining inter-individual differences in performance, aggregating over blocks of even 64 trials (panel **A**) can obscure meaningful variation. By instead fitting performance change over the course of the block (panel **B**), separable components of performance can be extracted. Here, the three measures of performance (panels **C**, **D**) would lead to divergent inferences about the relative order of the three participants. Figure adapted from Cochrane and Green (2021b)

correlated with all components of learning, while working memory ability was correlated only with asymptotic perceptual ability and had a vanishingly small correlation with the rate of learning.

Inter-individual differences in time-varying components are also useful in research contexts where learning is not the primary concern (Clarke & Hunt, 2023). For example, they can provide insights into the relationship between working memory and fluid intelligence (Cochrane & Green, 2021b; see also Fig. 4). Additionally, inferences about task performance can be made more theoretically informative by examining rapid changes in implicit social bias (Cochrane, Cox, et al., 2023a, 2023b), cognitive control (Cochrane et al., 2021), working memory encoding (Cochrane & Green, 2023), or other domains.

Beyond initial learning, we have limited knowledge of between-person variations that may constrain or facilitate perceptual generalization (Yang et al., 2020). However, the temporal specificity of by-trial and by-participant models enables the testing of specific components of generalization trajectories. Dale et al. (2021) found correlations between two visual perceptual learning tasks regarding the extent of initial generalization and the rate of generalization, even when controlling for initial training performance. They also showed links between the rate of generalization and response time measures of task switching (Rogers & Monsell, 1995), and between initial generalization and survey measures of individual differences in persistence, among other patterns of inter-individual variation.

Discussion

Here, we aimed to provide an overview of continuous-time estimates of performance and their applications in enhancing our understanding of learning processes. Continuous-time models utilize the entire time course of behavioral measurements to mitigate the influence of noise on performance estimates, thereby allowing for more robust analysis of learning dynamics. Estimated model parameters offer insights into various aspects of learning, including functional forms of change (and in turn the types of mechanisms or algorithms that could underpin the learning), generalization of learning, and individual differences in performance trajectories.

However, despite the numerous benefits offered by these approaches, there is still room for improvement and refinement. Specific parameterizations of change dynamics may not always be suitable, and researchers may encounter theoretical or practical challenges when implementing these types of models. It is important to acknowledge these limitations and explore avenues for further advancement and nuance in continuous-time modeling methodologies.

Beyond Improvements Due to Learning: Estimating Non-parametric Changes

Explicit parameterizations of change provide powerful tools for studying learning; not only are they more parsimonious and theory-driven than some other approaches (e.g., dividing trials into "blocks" and fitting several parameters to each block), comparing parameterizations can adjudicate between proposed mechanistic bases for learning (Cochrane et al., 2023c; Dosher & Lu, 2007; Heathcote et al., 2000). However, the utility of these parameterizations is realized only when their underlying assumptions are valid. For instance, exponential decay assumes a monotonic change in performance with a constant hazard function. However, in other cases of time series of behavioral performance, such assumptions may not be justified. The underlying function of change may be very complex (e.g., following several alternating periods of plateaus and improvements; see Cochrane et al., 2023c; K. M. Newell et al., 2009; Petrov et al., 2005; Yang et al., 2022). Alternatively, processes such as mind-wandering, strategy changes, or fatigue may be of theoretical interest. In such cases, it may be better to draw inferences from learning using nonparametric models.

Series of performance leading to learning can, in principle, be estimated using nonparametric approaches extending that described above (i.e., Bayesian mixed-effects models). Such modeling approaches may use blocks with smaller number of trials (Lu et al., 2023), basis functions or additive models over the dimension of time to approximate functions of the learning curve (Cochrane et al., 2023a), or latent state models (Ashwood et al., 2022). Other approaches are possible in principle (e.g., estimating autoregressive structures) but have not been investigated systematically, to our knowledge.

The use of basis functions to estimate changes over time can be implemented, like the above-described parametric approaches, for any given outcome variable (e.g., accuracy, psychometric function threshold, drift diffusion model drift rate). Given a vector of time values (e.g., trial numbers), a set of bases can be constructed (see Fig. 5 and Cochrane et al., 2023a). These bases can then be used to estimate the offset of that particular trial's performance (i.e., controlling for all other variables, how does that trial differ from other trials). Such an approach provides for the estimation of by-trial changes as well as improving the power to detect fixed effects (Cochrane et al., 2023a). The simplest form of this approach, using pre-defined radial bases with evenly-spaced centers and equal widths, can then be incorporated into any given mixed-effects model using conventional random-effects structures (i.e., using by-participant "slopes" to estimated each basis' coefficient). More complex approaches can

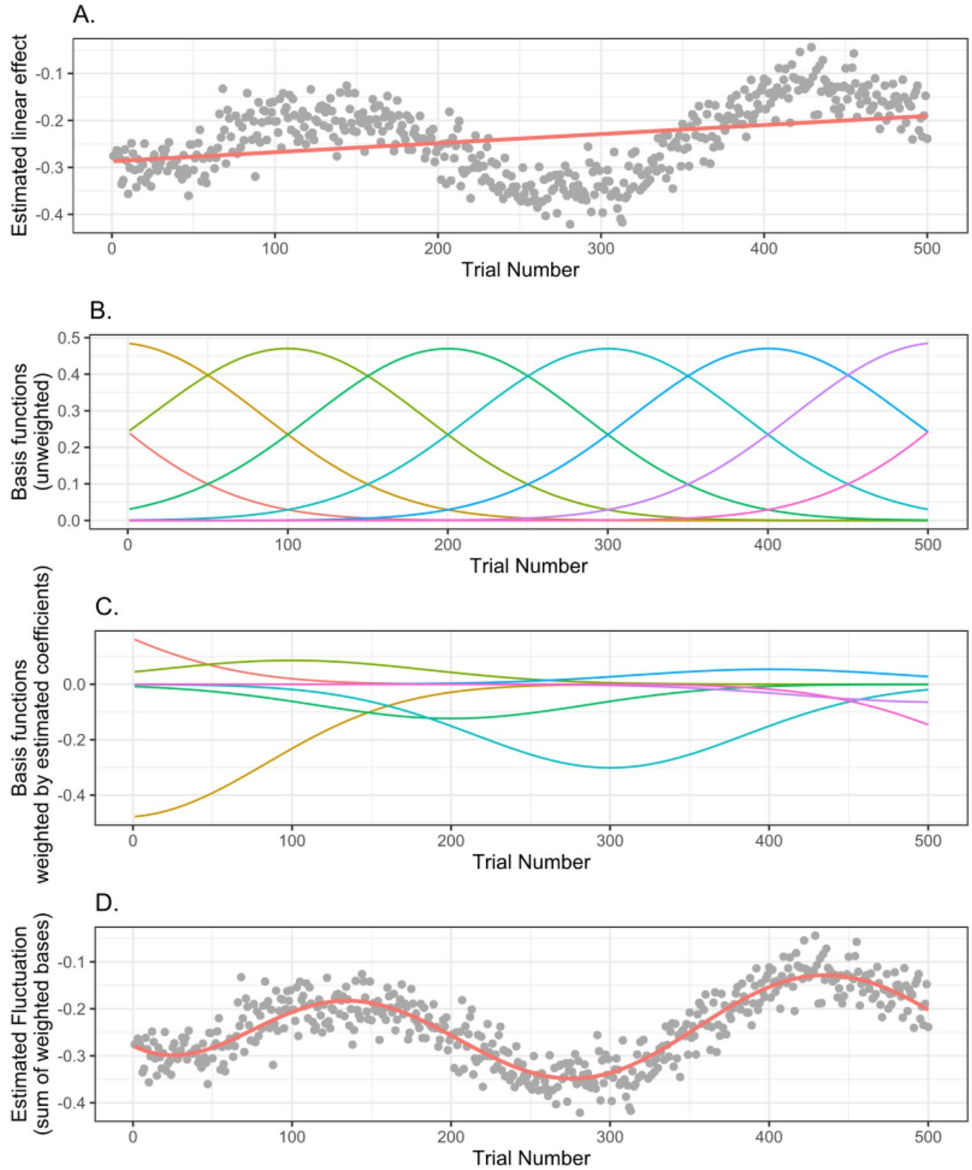

Fig. 5 A radial basis function approach to estimating fluctuations over time. Simulated data and its associated linear effect of time is shown in panel **A**, while panel **B** demonstrates the conversion of linear time (integers 1 through 500) into a set of overlapping unweighted basis functions. Panel **C** demonstrates how coefficient weights for these basis functions can be empirically estimated within a typical (i.e., [generalized] linear) regression model, thereby leading to the fitting of data-driven fluctuations in the measure of interest (panel **D**)

provide improvements in recovery of true fluctuations by estimating *b*-splines over the dimension of time, thereby creating a data-driven vector of time offset coefficients.

The HBM provides another potential non-parametric approach. In this approach, the thresholds in mini-blocks (typically 5–10 trials) are modeled as random variables that co-vary across subjects and time. It is very difficult to estimate a threshold from the few trials in each mini-block. However, the constraints from the between- and within-subject covariances and the conditional dependencies across levels of the hierarchical model could enable accurate and precise estimates of the thresholds in the mini-blocks (Zhao et al., xxxx).

We believe that our core proposition should be uncontroversial—aligning inferential methods with theoretical generative processes. Nonetheless, we recognize that various practical and theoretical difficulties can arise when seeking this alignment.

Practical Issues and Suggestions

When drawing inferences from the data generated by learning, it can be difficult to determine an appropriate method for aligning analysis and theory. This challenge may be due to the sparseness of data, complexity of experimental design, a large number of covariates, or other concerns. Fundamentally, however, we propose that existing methods (e.g., linear mixed-effects models) need relatively little augmentation to allow for testing trial-to-trial changes (Cochrane, 2020; Kattner et al., 2017a, 2017b). From this general-purpose baseline level (i.e., augmented [generalized mixed-effects] regression), a great amount of additional flexibility and power can be used to test hypotheses about time-dependent processes.

At a more mechanistic level, aligning inferences with generative processes must consider experimental design. For instance, if inferences are desired for changes occurring at a timescale smaller than blocks of trials, methods that only provide a point estimate (e.g., staircase methods leading to a single threshold estimate per run) would be inadequate. More generally, in line with other concerns in perceptual and cognitive training (Green et al., 2019), ensuring the utility and comparability of trials' data is desirable. Some standard methods may introduce inefficiencies or biases. Staircase runs typically include initial trials that are minimally informative regarding performance, and their conflation of experimental manipulation (e.g., stimulus strength) and performance (e.g., accuracy) restricts inferences. Alternately, using the method of constant stimuli is likely to mean that early trials may be uninformative due to being too difficult, while later trials may be uninformative due to being too easy. That is, because ability is improving while the distribution of stimulus strengths remains constant, stimulus strengths are likely to skew too-difficult early in learning and skew too-easy later; each of these skews reduces the informativeness of the trial's data for inferences regarding learning. These challenges can be addressed through appropriate adaptive methods, such as qCD (Zhang et al., 2019) or other approaches that balance trial informativeness and the ability to extrapolate beyond single-point estimates.

One additional implication of the framework put forward here is that perceptual and cognitive learning should not be categorized as having "non-learning trials" during pre-test or post-test; we consider it possible that learning, in some form, is occurring during all trials. Furthermore, we suggest that "learning trials" are at least as interpretable as "non-learning trials," if not more. We have reviewed methods for modeling time-varying dynamics in both parametric and non-parametric approaches, which facilitate the interpretation of both "pre-/post- test" and "training" tasks as learning environments. We presented existing methods and proposed a simple criterion for novel advances: Analytical models of learning should be able to make predictions of performance at a trial-by-trial level. This criterion carries several corollaries, explored in examples here, such as the need to be tested using trial-by-trial data and the need to be compatible with theories of performance improvement (e.g., Logan, 1988; A. Newell & Rosenbloom, 1981).

Theoretical Issues and Suggestions

A fundamental theoretical point we wish to emphasize is the importance of making the underlying assumptions more explicit when choosing an analytical method to draw inferences from data. In many cases, experimental methods used to test specific hypotheses are carefully designed and well-considered. Therefore, it is equally important to thoughtfully consider the inferential approach employed.

Specifically, we encourage researchers to reflect on how they handle their data concerning the assumptions of process stationarity (Molenaar & Campbell, 2009). When, for example, a particular index of performance is used to represent a certain number of trials (e.g., one single "threshold" for a block of 100 trials), this assumes both stationarity (the expectation that the generative process produces this threshold, on average, on each trial in the block) and ergodicity (treating variation over time as equivalent to variation around a single point estimate at a specific moment in time; Molenaar, 2008).

If a researcher intends to adopt these theoretical assumptions, this may be fully justified. However, we encourage greater explicitness about how aggregation, such as through averaging or fitting psychometric functions, implements these assumptions about the generative process. If such assumptions are not desired by the researcher, we hope the methods we have presented here provide the desired additional flexibility. The clarification of these assumptions should assist in making theories under investigation more transparent, as well as the evidence needed to differentiate between them.

These concerns regarding assumptions apply not only to learning but also to various processes that may not be stationary over time. Processes like mind-wandering, fatigue, exploring strategies, or other time-varying processes may be of theoretical importance and should be approached with continuous-time methods. Depending on the specific hypothesis or process, a continuous-time method may be best implemented as a model of parameterized change (e.g., monotonic adaptation or fatigue) or as a non-parametric model to uncover fluctuations in patterns that were not predefined (e.g., mind-wandering).

Considerations of temporal variations may be particularly important when studying special populations, as they might demonstrate greater-than-expected variability in processes of interest due to confounding factors such as distraction. By characterizing performance in a continuous-time way, such fluctuations can be estimated as well as controlled for.

Conclusions

Much like the shift observed over recent decades towards the adoption of mixed-effects models for a more comprehensive understanding of experimental data (e.g., Baayen et al., 2008; Barr et al., 2013; Lindstrom & Bates, 1990; Rouder & Lu, 2005), our present proposal is fundamentally centered on the alignment of inferential and theoretical perspectives. In the learning domains, especially, theorized mechanisms of change are inherently linked to predictions about the specific patterns that such changes should exhibit (Dosher & Lu, 2007; Karni & Sagi, 1993; Logan, 1988). In this context, we have demonstrated several ways in which the utilization of such theory-driven trial-to-trial change patterns can greatly enhance the investigation of many facets of perceptual and cognitive training.

Author Contribution A.C. wrote the initial draft of the manuscript. All authors contributed sections to the manuscript. All authors revised the manuscript and approved of the submitted version.

Funding This work was funded in part by grants N00014-22–1-2283 (Office of Naval Research) to CSG and EY017491 (NIH) to ZLL.

Data Availability No empirical data was collected or re-analyzed in the writing of this review.

Declarations

Competing Interests ZLL holds intellectual property interests in visual function measurement and rehabilitation technologies, and equity interests in Adaptive Sensory Technology, Inc. (San Diego, CA, USA) and Jiangsu Juehua Medical Technology Ltd (Jiangsu, China). CSG is the Editor-in-Chief of the Journal of Cognitive Enhancement.

References

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature, 387*(6631), 401–406. https://doi.org/10.1038/387401a0

Ashwood, Z. C., Roy, N. A., Stone, I. R., International Brain Laboratory, Urai, A. E., Churchland, A. K., Pouget, A., & Pillow, J. W. (2022). Mice alternate between discrete strategies during perceptual decision-making. *Nature Neuroscience, 25*(2), 201–212. https://doi.org/10.1038/s41593-021-01007-z

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research, 27*(6), 953–965. https://doi.org/10.1016/0042-6989(87)90011-3

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bejjanki, V. R., Beck, J. M., Lu, Z.-L., & Pouget, A. (2011). Perceptual learning as improved probabilistic inference in early sensory areas. *Nature Neuroscience, 14*(5), 642–648. https://doi.org/10.1038/nn.2796

Bejjanki, V. R., Zhang, R., Li, R., Pouget, A., Green, C. S., Lu, Z.-L., & Bavelier, D. (2014). Action video game play facilitates the development of better perceptual templates. *Proceedings of the National Academy of Sciences, 111*(47), 16961–16966. https://doi.org/10.1073/pnas.1417056111

Brown, S., & Heathcote, A. (2003). Averaging learning curves across and within participants. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society Inc, 35*(1), 11–21.

Bürkner, P. C. (2017). brms: An *R* Package for bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Clarke, A. D. F., & Hunt, A. R. (2023). *Learn more from your data with asymptotic regression* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/fkbza

Cochrane, A. (2020). TEfits: Nonlinear regression for time-evolving indices. *Journal of Open Source Software*, *5*(52), 2535. https://doi.org/10.21105/joss.02535

Cochrane, A., & Green, C. S. (2021a). Assessing the functions underlying learning using by-trial and by-participant models: Evidence from two visual perceptual learning paradigms. *Journal of Vision, 21*(13), 5. https://doi.org/10.1167/jov.21.13.5

Cochrane, A., & Green, C. S. (2021b). Trajectories of performance change indicate multiple dissociable links between working memory and fluid intelligence. *Npj Science of Learning, 6*(1), 33. https://doi.org/10.1038/s41539-021-00111-w

Cochrane, A., & Green, C. S. (2023). Working memory is supported by learning to represent items as actions. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-023-02654-z

Cochrane, A., Simmering, V., & Green, C. S. (2021). Modulation of compatibility effects in response to experience: Two tests of initial and sequential learning. *Attention, Perception & Psychophysics, 83*(2), 837–852. https://doi.org/10.3758/s13414-020-02181-1

Cochrane, A., Ruba, A. L., Lovely, A., Kane-Grade, F. E., Duerst, A., & Pollak, S. D. (2022). Perceptual learning is robust to manipulations of valence and arousal in childhood and adulthood. *PLoS ONE, 17*(4), e0266258. https://doi.org/10.1371/journal.pone.0266258

Cochrane, A., Cheng, Y.-A., & Watanabe, T. (2023a). *A regression-based framework for estimating and controlling for trial-to-trial variations in behavioral measurements* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/3z627

Cochrane, A., Cox, W. T. L., & Green, C. S. (2023b). Robust within-session modulations of IAT scores may reveal novel dynamics of rapid change. *Scientific Reports, 13*(1), 16247. https://doi.org/10.1038/s41598-023-43370-w

Cochrane, A., Sims, C. R., Bejjanki, V. R., Green, C. S., & Bavelier, D. (2023c). Multiple timescales of learning indicated by changes in evidence-accumulation processes during perceptual decision-making. *NPJ Science of Learning, 8*(1), 19. https://doi.org/10.1038/s41539-023-00168-9

Crossman, E. R. F. W. (1959). A theory of the acquisition of speed-skill. *Ergonomics, 2*(2), 153–166. https://doi.org/10.1080/00140135908930419

Dale, G., Sampers, D., Loo, S., & Green, C. S. (2018). Individual differences in exploration and persistence: Grit and beliefs about ability and reward. *PLoS ONE, 13*(9), e0203131. https://doi.org/10.1371/journal.pone.0203131

Dale, G., Cochrane, A., & Green, C. S. (2021). Individual difference predictors of learning and generalization in perceptual learning. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-021-02268-3

Dayan, P., & Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience, 8*(4), 429–453. https://doi.org/10.3758/CABN.8.4.429

Deveau, J., Lovcik, G., & Seitz, A. (2013). The therapeutic benefits of perceptual learning. *Current Trends in Neurology, 7*, 39–49.

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. https://doi.org/10.1111/j.1467-9280.2007.01934.x

Dosher, B., & Lu, Z.-L. (2020). Perceptual learning: How experience shapes visual perception. The MIT Press.

Eckhoff, P., Holmes, P., Law, C., Connolly, P. M., & Gold, J. I. (2008). On diffusion processes with variable drift rates as models for decision making during learning. *New Journal of Physics, 10*, nihpa49499. https://doi.org/10.1088/1367-2630/10/1/015006

Fahle, M. (2005). Perceptual learning: Specificity versus generalization. *Current Opinion in Neurobiology, 15*(2), 154–160. https://doi.org/10.1016/j.conb.2005.03.010

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. https://doi.org/10.1016/0042-6989(93)90094-D

Gallistel, C. R., Fairhurst, S., & Balsam, P. (2004). The learning curve: Implications of a quantitative analysis. *Proceedings of the National Academy of Sciences, 101*(36), 13124–13131. https://doi.org/10.1073/pnas.0404965101

Gold, J. I., & Ding, L. (2013). How mechanisms of perceptual decision-making affect the psychometric function. *Progress in Neurobiology, 103*, 98–114. https://doi.org/10.1016/j.pneurobio.2012.05.008

Green, C. S., Pouget, A., & Bavelier, D. (2010). Improved probabilistic inference as a general learning mechanism with action video games. *Current Biology, 20*(17):1573–1579. https://doi.org/10.1016/j.cub.2010.07.040

Green, C. S., Bavelier, D., Kramer, A. F., Vinogradov, S., Ansorge, U., Ball, K. K., Bingel, U., Chein, J. M., Colzato, L. S., Edwards, J. D., Facoetti, A., Gazzaley, A., Gathercole, S. E., Ghisletta, P., Gori, S., Granic, I., Hillman, C. H., Hommel, B., Jaeggi, S. M., & Witt, C. M. (2019). Improving methodological standards in behavioral interventions for cognitive enhancement. *Journal of Cognitive Enhancement, 3*(1), 2–29. https://doi.org/10.1007/s41465-018-0115-y

Harlow, H. F. (1949). The formation of learning sets. *Psychological Review, 56*(1), 51–65.

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.

Fiorentini, A., & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature, 287*(5777), 43–44. https://doi.org/10.1038/287043a0

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. https://doi.org/10.1167/9.3.1

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. https://doi.org/10.1016/j.visres.2010.06.016

Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature, 365*(6443), 250–252. https://doi.org/10.1038/365250a0

Kattner, F., Cochrane, A., Cox, C. R., Gorman, T. E., & Green, C. S. (2017a). Perceptual learning generalization from sequential perceptual training as a change in learning rate. *Current Biology, 27*(6), 840–846. https://doi.org/10.1016/j.cub.2017.01.046

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

Kruschke, J. K. (2011). *Doing bayesian data analysis: A tutorial with R and BUGS*. Academic Press.

Kruschke, J. K., & Vanpaemel, W. (2015). *Bayesian estimation in hierarchical models* (J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels, Eds.; Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199957996.013.13

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology, 55*(1), 1–7. https://doi.org/10.1016/j.jmp.2010.08.013

Lesmes, L. A. (2010). Bayesian adaptive estimation of the contrast sensitivity function: The quick CSF method. *Journal of Vision, 10*(3), 1–21. https://doi.org/10.1167/10.3.17

Lesmes, L. A., Lu, Z.-L., Baek, J., Tran, N., Dosher, B. A., & Albright, T. D. (2015). Developing Bayesian adaptive methods for estimating sensitivity thresholds (d′) in Yes-No and forced-choice tasks. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01070

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. https://doi.org/10.1016/j.visres.2006.04.022

Lindstrom, M. J., & Bates, D. M. (1990). Nonlinear Mixed Effects Models for Repeated Measures Data. *Biometrics, 46*(3), 673. https://doi.org/10.2307/2532087

Liu, J., Lu, Z.-L., & Dosher, B. A. (2012). Mixed training at high and low accuracy levels leads to perceptual learning without feedback. *Vision Research, 61*, 15–24. https://doi.org/10.1016/j.visres.2011.12.002

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492–527.

Lu, Z.-L., Zhao, Y., Liu, J., & Dosher, B. (2023). Non-parametric hierarchical Bayesian Modeling of the learning curve in perceptual learning. *Journal of Vision, 23*(9), 5752. https://doi.org/10.1167/jov.23.9.5752

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research & Perspective, 2*(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

Molenaar, P. C. M. (2008). On the implications of the classical ergodic theorems: Analysis of developmental processes has to focus on intra-individual variation. *Developmental Psychobiology, 50*(1), 60–69. https://doi.org/10.1002/dev.20262

Molenaar, P. C. M., & Campbell, C. G. (2009). The New Person-Specific Paradigm in Psychology. *Current Directions in Psychological Science, 18*(2), 112–117. https://doi.org/10.1111/j.1467-8721.2009.01619.x

Montague, P. R., & Sejnowski, T. J. (1994). The predictive brain: Temporal coincidence and temporal order in synaptic learning mechanisms. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *1*(1), 1–33.

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). Lawrence Erlbaum.

Newell, K. M., Mayer-Kress, G., Hong, S. L., & Liu, Y.-T. (2009). Adaptation and learning: Characteristic time scales of performance dynamics. *Human Movement Science, 28*(6), 655–687. https://doi.org/10.1016/j.humov.2009.07.001

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. https://doi.org/10.1037/0033-295X.112.4.715

Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*(2), 207–231. https://doi.org/10.1037/0096-3445.124.2.207

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review, 12*(4), 573–604. https://doi.org/10.3758/BF03196750

Rouder, J. N., Sun, D., Speckman, P. L., Lu, J., & Zhou, D. (2003). A hierarchical bayesian statistical framework for response time distributions. *Psychometrika, 68*(4), 589–606. https://doi.org/10.1007/BF02295614

Seitz, A., Kim, D., & Watanabe, T. (2009). Rewards evoke learning of unconsciously processed visual stimuli in adult humans. *Neuron, 61*(5), 700–707. https://doi.org/10.1016/j.neuron.2009.01.016

Seitz, A., Nanez, J. E., Holloway, S., Tsushima, Y., & Watanabe, T. (2006). Two cases requiring external reinforcement in perceptual learning. *Journal of Vision, 6*(9), 966–973. https://doi.org/10.1167/6.9.9

Shiu, L., & Pashler, H. (1994). Negligible effect of spatial precuing on identification of single digits. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1037–1054.

Snoddy, G. S. (1926). Learning and stability: A psychophysiological analysis of a case of motor learning with clinical applications. *Journal of Applied Psychology, 10*(1), 1–36. https://doi.org/10.1037/h0075814

Song, M., Behmanesh, I., Moaveni, B., & Papadimitriou, C. (2020). Accounting for modeling errors and inherent structural variability through a hierarchical Bayesian model updating approach: An overview. *Sensors (basel, Switzerland), 20*(14), 3874. https://doi.org/10.3390/s20143874

Watanabe, T., Náñez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature, 413*(6858), 844–848. https://doi.org/10.1038/35101601

Watson, A. B., & Pelli, D. G. (1983). Quest: A Bayesian adaptive psychometric method. *Perception & Psychophysics, 33*(2), 113–120. https://doi.org/10.3758/BF03202828

Wilson, J. D., Cranmer, S., & Lu, Z.-L. (2020). A hierarchical latent space network model for population studies of functional connectivity. *Computational Brain & Behavior, 3*(4), 384–399. https://doi.org/10.1007/s42113-020-00080-0

Wimer, R. (1964). Osgood's transfer surface: Extension and test. *Journal of Verbal Learning and Verbal Behavior, 3*(4), 274–279. https://doi.org/10.1016/S0022-5371(64)80066-9

Woodworth, R. S., & Thorndike, E. L. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review, 8*(3), 247–261. https://doi.org/10.1037/h0074898

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. https://doi.org/10.1016/j.cub.2008.10.030

Yang, J., Yan, F.-F., Chen, L., Xi, J., Fan, S., Zhang, P., Lu, Z.-L., & Huang, C.-B. (2020). General learning ability in perceptual learning. *Proceedings of the National Academy of Sciences, 117*(32), 19092–19100. https://doi.org/10.1073/pnas.2002903117

Yang, J., Yan, F.-F., Chen, L., Fan, S., Wu, Y., Jiang, L., Xi, J., Zhao, J., Zhang, Y., Lu, Z.-L., & Huang, C.-B. (2022). Identifying Long- and Short-Term Processes in Perceptual Learning. *Psychological Science, 33*(5), 830–843. https://doi.org/10.1177/09567976211056620

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

Zhang, R.-Y., Chopin, A., Shibata, K., Lu, Z.-L., Jaeggi, S. M., Buschkuehl, M., Green, C. S., & Bavelier, D. (2021). Action video game play facilitates "learning to learn." *Communications Biology, 4*(1), 1154. https://doi.org/10.1038/s42003-021-02652-7

Zhao, Y., Lesmes, L. A., Dorr, M., & Lu, Z.-L. (2021). Quantifying uncertainty of the estimated visual acuity behavioral function with hierarchical Bayesian modeling. *Translational Vision Science & Technology, 10*(12), 18. https://doi.org/10.1167/tvst.10.12.18

Zhao, Y., Lesmes, L., & Lu, Z.-L. (2019). Efficient assessment of the time course of perceptual sensitivity change. *Vision Research, 154*, 21–43. https://doi.org/10.1016/j.visres.2018.10.009

Zhao, Y., Liu, J., Dosher, B. A., & Lu, Z.-L. (submitted). *Identifying component processes in perceptual learning with non-parametric hierarchical Bayesian modeling of the learning curve*.