Journal of Experimental Psychology: General

Learn More From Your Data With Asymptotic Regression

Alasdair D. F. Clarke and Amelia R. Hunt

Online First Publication, February 17, 2025. https://dx.doi.org/10.1037/xge0001710

AMERICAN PSYCHOLOGICAL ASSOCIATION
Journal of Experimental Psychology: General

ISSN: 0096-3445
https://doi.org/10.1037/xge0001710

Learn More From Your Data With Asymptotic Regression

Alasdair D. F. Clarke[1] and Amelia R. Hunt[2]
[1] Department of Psychology, University of Essex
[2] School of Psychology, University of Aberdeen

All measures of behavior have a temporal context. Changes in behavior over time often take a similar form: monotonically decreasing or increasing toward an asymptote. Whether these behavioral dynamics are the object of study or a nuisance variable, their inclusion in models of data makes conclusions more complete, robust, and well-specified, and can contribute to theory development. Here, we demonstrate that asymptotic regression is a relatively simple tool that can be applied to repeated-measures data to estimate three parameters: starting point, rate of change, and asymptote. Each of these parameters has a meaningful interpretation in terms of ecological validity, behavioral dynamics, and performance limits, respectively. They can also be used to help decide how many trials to include in an experiment and as a principled approach to reducing noise in data. We demonstrate the broad utility of asymptotic regression for modeling the effect of the passage of time within a single trial and for changes over trials of an experiment, using two existing data sets and a set of new visual search data. An important limit of asymptotic regression is that it cannot be applied to data that are stationary or change nonmonotonically. But for data that have performance changes that progress steadily toward an asymptote, as many behavioral measures do, it is a simple and powerful tool for describing those changes.

Public Significance Statement
Experiments in psychology often involve measuring behavior repeatedly over time. Sometimes understanding the way behavior changes over time is the goal of the experiment, but often it is not. Even if they are not of direct interest, changes in behavior over time make average measurements difficult to interpret or compare across conditions. In this article, we point out that changes over time often take a similar form that can be statistically modeled using asymptotic regression. We present three applications that demonstrate the range of problems this approach can solve.

Keywords: learning, modeling, timecourse, repeated-measures

When measuring the effect of an independent variable or manipulation on behavior, we have to measure it in the context of many other factors that influence our measurements, some of which we can control, and some of which we cannot. For example, there will be differences between individual participants in a given sample that we cannot control and may not be relevant to our research question, but add substantial variability to the data. The standard statistical approach to dealing with individual differences in behavioral measurements is to estimate their effect on the data and remove them from the statistical model (e.g., by applying a repeated-measures analysis of variance or including participant as a random factor in a multilevel model). This approach is so routine when it comes to differences between individuals that it is a fundamental part of most introductory undergraduate statistics courses in psychology. There

Timothy Vickery served as action editor.
All data and analysis code are available on a public Open Science Framework repository (https://osf.io/64r7m/). The first draft was uploaded to PsyArXiV on October 20, 2023 (https://osf.io/preprints/psyarxiv/fkbza). The ideas in this article were presented in part at the annual meeting of the Vision Sciences Society in May 2023 and at the European Conference on Visual Perception in August 2024.
This project was supported by an Economic and Social Research Council standard grant (ES/S016120/1) with Amelia R. Hunt as principal investigator and Alasdair D. F. Clarke as coinvestigator. The authors are grateful to Aaron Cochrane for comments on the draft and to the reviewers for comments on the submitted article. We are also grateful to Sitong Lin for her contributions to Example 3.

Alasdair D. F. Clarke played a lead role in data curation, formal analysis, investigation, methodology, project administration, software, and visualization, a supporting role in writing–original draft, and an equal role in conceptualization and writing–review and editing. Amelia R. Hunt played a lead role in resources and writing–original draft and an equal role in writing–review and editing.
Correspondence concerning this article should be addressed to Alasdair D. F. Clarke, Department of Psychology, University of Essex, Colchester CO4 3SQ, United Kingdom. Email: a.clarke@essex.ac.uk

are of course also many situations in which individual differences are part of the theory being tested, in which case the random effect of the participant is not (or should not be) discarded, but included as part of the statistical model of the data used to test hypotheses. But whether a researcher is interested in individual differences or not, the logic of estimating their contribution to variation in the data is well-accepted and broadly applied. By attributing variance to known sources, we boost our sensitivity to measure the effects of the independent variable we are interested in.

An equally ubiquitous variable in experimental psychology is time, but the conventions around estimating the effects of time are far less established than for the effects of the participant. When behavior is relatively stable over time, a central tendency measure is a simple and meaningful way to aggregate a lot of information into a single estimate. But when there are substantial changes over time, the median or mean performance might be an accurate representation of behavior only at a single timepoint in the trial sequence, and presenting this piece of information outside of its temporal context is to commit what Alfred North Whitehead would call the "Fallacy of Misplaced Concreteness" (1929). When behavior is in flux, the unit of analysis cannot be a single number. A similar principle applies for time as applies to participant differences; whether time is a theoretically meaningful variable in your experiment or not, accounting for the effects of time on the data means estimating and attributing variability to time that would otherwise be considered random noise.

Many classic psychological phenomena are embedded within the context of a short-term flux, as behaviors emerge or dissipate within the span of a single trial. Some effects, such as adaptation, take some time to develop to their optimum (Newell et al., 2009). Speed–accuracy trade-offs are ubiquitous as well, with slower responses typically (but not always) associated with decreasing error rates and higher motor precision. In specific subfields, theories of the underlying mechanisms of temporal patterns lead to specific predictions about how changes unfold within a trial. For example, the timecourse of visual selection can be observed in a phenomenon known as oculomotor capture (Theeuwes et al., 1998), in which the tendency for eye movements to inadvertently deviate toward visually salient distractors diminishes as the latency of the saccade increases. In task-switching, the cost of switching between tasks tends to get smaller with longer preparatory intervals, suggesting participants are able to engage in active, time-consuming reconfiguration of their readiness to perform a specific task (Koch, 2003). In perceptual aftereffects and other serial dependencies, the size of the bias induced by the preceding event diminishes with time (e.g., Manassi et al., 2023). Across these diverse examples, the shape of the timecourse is similar, with effects, biases, costs, and errors monotonically either decreasing or increasing over time toward an asymptote.[1] This similarity provides an opportunity for a general-purpose solution that could more precisely represent the effects of interest, and ease comparison from one study to another.

Performance changes over time are also inherent in any series of repeated trials. Our usual approach to addressing the problem of variability due to time in our own research has been to follow the norm in experimental psychology: We include a set of practice trials. Practice trials tend to be excluded from further analysis, and sometimes not even recorded. We rarely scrutinize these practice trials to evaluate whether they were sufficient, to learn information from them, or to consider their potential to limit or extend generalizability. After practice, the rest of the trials are typically considered as a stationary set, with the unit of analysis being a measure of central tendency taken from all the trials within a given condition/participant. But even beyond the early learning of instructions and response mappings, as participants gradually experience all the conditions of the experiment, they can adapt their strategies to optimize performance, tune their attention to pick up information from the appropriate locations, and start to anticipate sequences of events and their timing with more precision. As responses become more automated, there may even be a shift from one underlying mechanism to another. Logan's classic instance theory of automaticity (Logan, 1988), for example, proposes that responses to previous encounters with the same stimulus are stored and retrieved when we encounter the same stimulus again. The lower limit on response time (RT) depends on how quickly the algorithm leading to a response can be executed, or on how quickly the memory of the response executed in the previous encounter with that stimulus can be retrieved and instantiated, whichever is fastest. As the number of instances builds in memory, the probability that memory retrieval drives the response increases. In other words, RTs at the start of an experiment may be determined by qualitatively different mental processes than those executed later, even though the task is the same. An average of the whole trial sequence will therefore represent some midway point between two mental processes, which is an unproductive approach to understanding either of them. Instead, both timepoints, as well as the transition between them, are more meaningful measures of human performance.

A fundamental part of the design of most behavioral experiments that is rarely questioned or discussed is how much practice is needed. Almost all experiments include some practice trials to ensure the participant understands the instructions, but exactly how many trials to include seems to be determined by an unspoken combination of convention, instinct, time constraints, and personal preference. Although the underlying purpose of practice is to stabilize performance, we rarely scrutinize whether the practice provided was sufficient to achieve this end. Being able to evaluate how much practice is sufficient can also ensure we are not wasting participants' limited attention span on unnecessary practice trials. To address this ubiquitous dilemma of experimental design, the same tools that can provide a description of the effects of time can also be applied to remove noise due to the effects of time when these effects are not of interest. The stable differences that emerge between controlled conditions or groups, irrespective of time, will be a cleaner estimate of these differences. The implication is that there need be no separation between "practice" and "experimental" trials in any experiment. Instead, we can use the full set of data to estimate the start and end points of behavior and the rate of change between them. This information can be useful in deciding how many trials are needed to achieve stable performance, addressing the question of not only how many practice trials are needed (i.e., none), but also how many experimental trials are needed (i.e.,

[1] The precise function the change follows is, in many specific areas of research, an important constraint on theory (e.g., Cochrane & Green, 2021; Heathcote & Brown, 2000). In this article, we are not comparing functions but instead presenting one simple function that usually provides a reasonably good fit for behavioral change and can be broadly applied.

enough that performance estimates no longer change much as more trials are added).

Perceptual and statistical learning are substantial subfields within psychology with the goal of understanding the mechanisms underlying changes in performance over a range of timescales and conditions. In perceptual learning (e.g., Goldstone, 1998), changes can happen gradually as participants learn and streamline procedures and transformations, or they can happen suddenly, like the classic examples of Mooney faces where there is an instantaneous and permanent change in how the visual information is organized and interpreted. In the literature on contextual cueing (e.g., Chun & Jiang, 1998) and statistical learning in a visual search context (e.g., Wang & Theeuwes, 2018), the focus is on the extent to which participants learn, over a series of repeated trials, to exploit nonrandom spatial and temporal features of the contexts to guide spatial selection, refine target templates, and suppress distractors. In these fields, change over time is often characterized blockwise, or with an arbitrary division of trials into quantiles within which performance is averaged and between which it is compared. While this is a useful approach for determining whether change occurred, and to some extent for measuring how large that change is, it does not allow for any interpretation of the rate of change itself, as noted in Kershner and Hollingworth (2023). Similarly, Kattner et al. (2017) have criticized the conventional approach to analyzing timecourse data in the perceptual learning literature, whereby psychometric functions are fit to single estimates summarizing a series of blocks, each containing many trials, ignoring the likely case that there are changes within blocks as well as between them. They advocate for modeling the trial-dependent changes that are a close match to the continuous learning that presumably underlies these changes over time. We agree with this point and argue here that modeling time-dependent changes does not need to be restricted to situations where these changes over time are the subject of investigation. Rather, we should treat time the way we treat differences between participants, routinely estimating variability due to time and either including it as part of our statistical models (if it is relevant to our predictions) or removing it from the category of "unexplained noise" (if it is not relevant).

In summary, changes over time are a fundamental part of human behavior, and these often take the same form: a monotonic increase or decrease toward an asymptote. Some researchers are developing theories that make predictions about time effects, but lack a simple convention that can be applied across many circumstances to estimate the span and rate of change. Other researchers are not interested in the effects of time per se, but could be using them to design more efficient experiments and to provide a more refined estimate of effects of interest. Either way, researchers who are measuring human behavior at more than a single timepoint will benefit from converging on a set of conventions about how to account for the effects of time in their experiments, and we propose a simple solution here and provide four diverse examples of how to apply it.

Asymptotic Regression

Asymptotic regression refers to a class of nonlinear models in which the predicted value y approaches some asymptote as the predictor x approaches infinity. Perhaps the simplest of these models (Stevens, 1951) is defined by the exponential equation:

$$\mu = a - (a - b)\rho^t, \quad (1)$$

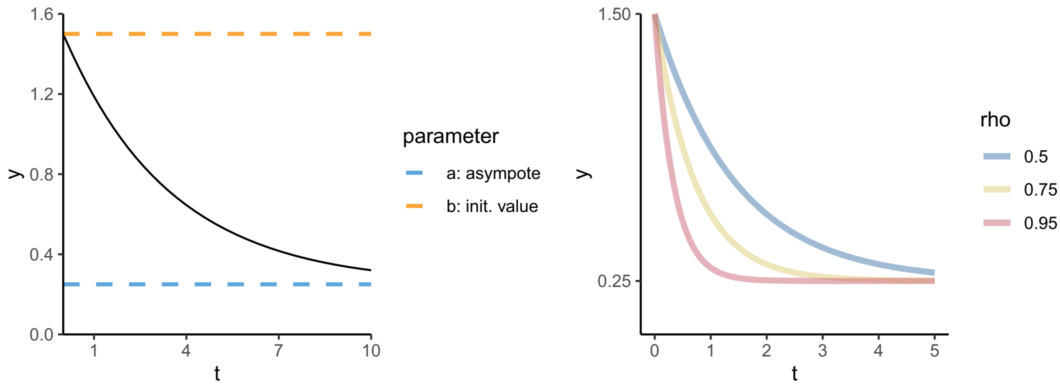
where $\rho \in (0, 1)$. It is worth noting that is equivalent to $\mu = a - (a - b)e^{-st}$ where $s = -\log(\rho)$. The model outlined in Equation 1 is just one specific approach from a more general set of models that seek to deal with temporal dynamics. Depending on context, more sophisticated models such as splines, or those outlined by Speekenbrink and Shanks (2010), may offer a better description of these dynamics (e.g., see Pasqualotto et al., 2024). That said, many common patterns in behavioral data such as those described above are well-suited to asymptotic regression (practice effects, learning, performance gains with longer preparatory intervals or exposure durations), and the three parameters in the regression all have meaningful interpretations.

Parameter Interpretations

- The *initial value* (b) is behavior at a hypothetical "time 0." For issues of ecological validity, and generalizing to predict behavior in applied or new settings, this can be a useful estimate of what performance would have looked like before performance was altered over time or by repeated exposure. This parameter improves on just looking at early performance in isolation because it uses information from the whole time series to estimate the start point.
- The *rate* (ρ) describes how performance curves from the initial value toward the asymptote, and provides a single, simple measure of the effect of time on performance data. This is a more sensitive measure than simply comparing performance change from the start to the end of a period of time; two manipulations could lead to the same change overall, but if one produces that change more quickly than the other, this will be reflected in the rate. In order for us to obtain an asymptote, ρ must be $0 < \rho < 1$. We enforce this constraint by using a logistic transform, in much the same manner as one would when fitting generalized linear models with link functions.
- The *asymptote* (a) is the hypothetical projected performance as it flattens, which is useful when the researcher is chiefly interested in "best case" performance. This is often the case in a situation involving comparison to theoretical predictions and optimal performance, or with an AI. Asymptotes can also be used to compare between groups of participants who may differ in their familiarity with the laboratory setting or who take longer to learn and remember response mappings or task rules. They are also useful when effects are anticipated to be small relative to the noise of early trials and learning.

See Figure 1 for an illustration of the three parameters. All three taken together provide a more complete and accurate description of the flux of behavior that is present in many data sets. Of course, not all data take this form. As we return to in the discussion, there are other effects of time on behavior that require different tools. But when we started opportunistically sampling and examining data from our own work and others, we found asymptotic regression was often a good model, and when it was, it led to new insights about the

Figure 1
Illustration of the Three Model Parameters



Note. (Left) An example of the asymptotic regression curve with the initial value (b) and asymptote (a) highlighted. (Right) An example of three different values of ρ. See the online article for the color version of this figure.

data that could not have come from examining standard measures of central tendency. Despite the usefulness of this approach for describing human behavioral data, however, we could find few examples of its application in psychology outside of the learning literature (e.g., Cochrane et al., 2018; Cochrane & Green, 2021; Rast, 2011; Williams et al., 2019). A notable exception is Cochrane, Cox, and Green (2023), who used a similar approach to what we advocate here to explore how performance on a classic test of latent bias (the Implicit Association Test) changes over the course of trials and blocks of trials. This approach uncovers differences between individuals that go beyond a simple effect size between conditions, but instead reveal more distinct and specified differences in start point, rate, and asymptote (see also Cochrane, Sims, et al., 2023). Our aim in the current article, therefore, was to advocate for the broader use of asymptotic regression for modeling time series effects and to provide some simple worked examples of the additional insight that can be gained from using it. We selected these three examples somewhat arbitrarily, with the only constraint being that the experiments are relatively simple to describe and that they present distinct scenarios. We first provide a more precise description of the asymptotic regression model.

Model Equation

We can model the relationship between t (in this example, t is a trial, but it could also be other units of time) and our outcome (dependent variable) y as Equation 2:

$$\begin{aligned} y &\sim \mathcal{N}(\mu, \sigma^2) \\ \mu &= a - (a - b)(1 - \rho)^{t/t_s} \\ \text{logit}(\rho) &= r, \end{aligned} \quad (2)$$

where a (asymptote), b (initial value), r (i.e., the inverse logit of the rate), and σ are parameters to be estimated from the data. t is the trial number, and t_s is some standardizing constant that helps us avoid computational difficulties due to floor and ceiling effects (discussed below).[2] Some examples of this function are shown in Figure 1. We use $1 - \rho$ rather than simply ρ so that larger values lead to faster progress from the initial value to the asymptote (see Figure 1).

In most situations, this relationship is not our primary interest. Instead, we wish to analyze how behavior varies between experimental conditions, or with some numerical covariate. In these cases, we can allow a, b, and ρ to vary such measures (Equation 3):

$$\begin{aligned} a &\sim \beta_a X \\ b &\sim \beta_b X \\ r &\sim \beta_\rho X. \end{aligned} \quad (3)$$

After fitting such a model, we can examine the posterior distributions and characterize the extent to which predictors (experimental manipulations, independent variables) lead to changes in a, b, and ρ. We also recommend generating posterior predictive plots. The model can be extended to incorporate multilevels (i.e., random effects, hierarchical modeling) by adding a random effect structure to the linear models for a, b, and r.

Interpreting ρ and Selecting t_s

The asymptote and starting point, a and b, have straightforward interpretations: The values are measured in the same units as our dependent variable y. One point to note here is that the model estimates performance at a hypothetical point where $t = 0$. It is therefore important to consider what 0 means in the context of your data. For example, if your raw time unit is based on a computer's system time, or the time of day, 0 is meaningless and you will need to adjust your time units so that 0 reflects the start of the time series you are modeling.

Giving a meaning to ρ (the rate) is perhaps a little less intuitive. To simplify, consider a case where $t_s = 1$ and $a = 0$. In this case, $y = (1 - \rho)^t$, so we move $(1 - \rho)$ of the way toward 0 for every unit increase in t. In other words, if $\rho = 1/2$, y halves every time we go from x to $x + 1$. If a is some arbitrary value, then $1 - \rho$ is the proportion of the way we move from our current value to a with each

[2] We originally denoted this parameter as t_0, and it appears this way in some of our code. We changed to using t_s in the article to avoid any potential confusion with b, which represents starting point performance, that is, when $t = 0$.

increase in t. We have to complicate things slightly by including a normalizing constant t_s. This is due to the fact that t could be measured in a wide range of units such as seconds, milliseconds, or the tth trial in a block of trials. This means that the effect of moving from t to $t + 1$ could be wildly different depending on the time range and units. Our proposed solution is to normalize t by a constant t_s, in much the same way as z scores of a variable can be used.

This leads to the obvious question: How should t_s be selected? Our advice is to plot your dependent variable against your time range and set t_s to the value of t which lines up with the point at which y has moved *approximately* halfway from the initial value to the asymptote. This leads to estimates of $\rho \approx 1/2$ and $r \approx 0$. This facilitates easier computational estimating by avoiding extremely small or large values of r. Ideally, the choice of t_s could be made based on pilot data, or data from similar studies. There is no need to be overly precise in the choice of this hyperparameter; the aim is simply to avoid ρ ending up very close to 0 or 1.

Implementation Details

We have implemented this model in Stan (Stan Development Team, 2020), a probabilistic programming language for fitting Bayesian models.[3] As this model is somewhat more complex than a standard linear regression, care should be given to the choice of priors. We favor using weakly informative priors that guide the algorithm to plausible regions of the parameter space without overly biasing the model one way or another. For the random effect structure, we follow Lewandowski et al. (2009) and use an LKJ prior. Code for the full multilevel model is available at https://osf.io/64r7m/. For more about priors, see the Limitations section of the discussion.

Two of our examples involve modeling RTs. RT distributions are often strictly positive and highly skewed with a long tail. We have chosen to log-transform RTs in Example 2 to demonstrate the model with a simple approach to skewed distributions. In Example 3, we demonstrate a more sophisticated approach, using a lognormal distribution. In theory, asymptotic regression can be applied to more complex models of reaction time data, such as shifted-lognormal (Hughes et al., 2024), ex-Gaussian (Cochrane et al., 2021), and drift-diffusion models (Cochrane, Sims, et al., 2023), but our purpose here is to focus on the time series modeling, so we stick to simpler approaches.

In addition to asymptotic regression, we also fit linear ($\mu = bx + a$) and stationary ($\mu = a$) models in order to demonstrate the extent to which asymptotic regression can offer a closer fit to the data. We used LOO (leave-one-out) with Pareto smoothed importance sampling (Vehtari et al., 2024) to approximate how well each model can predict a held-out point when trained on the rest of the data. We then computed model weights to compare the three models. These tell us how often we should expect each model to outperform the other models under consideration. (Alternatively, we can interpret these as the weights with which we should combine all three models to give the best overall prediction.) One complication with this procedure is that a small (<1%) proportion of the data has unreliable Pareto k statistics. This appears to be mainly due to the initial few points being overly influential in the estimate of the b parameters.[4] To deal with this issue, we adopted an iterative process in which we repeatedly (a) fit the model, (b) calculate LOO Pareto k statistics, and (c) remove any points from the data with a Pareto k greater than 0.7. This process is repeated until we obtain LOO statistics with no warnings.

Simulations Illustrate the Problem With Ignoring Time

The standard approach to summarizing data over repeated measures is to use a measure of central tendency (e.g., mean[5]) and variance, typically with an assumption that the distribution is normal. In the additional online materials at https://osf.io/64r7m/, we have presented model comparisons when asymptotic regression, linear regression, and a normal distribution are applied to simulated data with different underlying trends. These demonstrate the general point that unless there are no trends in the data, the asymptotic model minimizes residual error, a point we illustrate with human data in the three more complex examples that we present in the rest of this article. There are problems with ignoring time that can extend beyond that of simply not accounting for variance, however. In many experiments in psychology, practice effects (e.g., slower RTs early in the experiment) will influence the mean, and the more trials there are in an experiment, the smaller that influence becomes. This can be seen in Figure 2. The beige interval shows the range of estimates of the mean of simulated data (the points) as the number of trials in the experiment increases from 1 to 500. The mean clearly drops as the number of trials increases. This creates a serious confound when comparing conditions that differ in the numbers of trials, the amount of practice, or the rate of learning: Any difference between means, or lack of difference, becomes impossible to interpret. In contrast, the estimate of the asymptote (the blue distribution in the plot) is uncertain when based only on the early trials, which is entirely appropriate. As the trial sequence progresses, the estimate quickly becomes more certain, and by about trial 100 (in this simulation), the asymptote estimate is more certain than the mean. Importantly, the asymptote parameter is not biased toward early trials like the mean, but remains unbiased. As a result, estimates of the asymptote from one condition can be compared to another condition with fewer trials, while estimates of the mean cannot.

Issues with comparing conditions with different numbers of trials have been raised before (e.g., Miller, 1988). A relevant related point is made nicely by Rouder and Haaf (2019), who note that effect sizes and reliability tend to increase as the number of “replicates,” or trials, increases. The problem they point out is that this leads to a lack of *portability*. Portability is an underappreciated standard we should be striving for in our measurements, that effect size estimates should be reliable, that is, stable from one sample to another. This allows researchers to state (within some confidence interval) how large an effect actually is, which is important for practical applications as well as for being able to plan future experiments with appropriate statistical power. Psychological tasks often fail to meet this portability standard because from one experiment or condition or population to

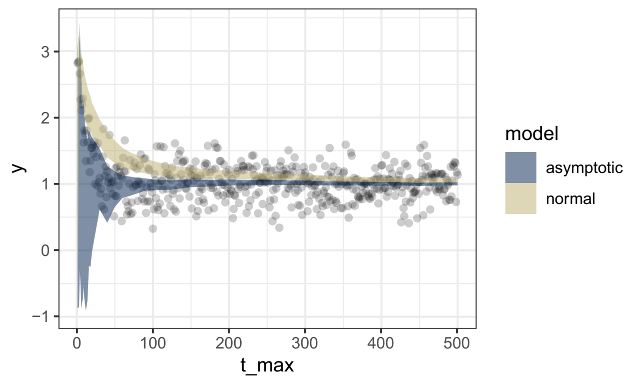
[3] There is nothing inherently “Bayesian” about these models. They could be fit using frequentist methods and maximum likelihood estimation.

[4] Pareto k statistics can be thought of as a Bayesian equivalent to Cook’s distance. These points are not necessarily being poorly fit by the model. Instead, they are an indication that they are particularly important in determining which set of parameter values best fits the data.

[5] We discuss the problem in relation to comparing means because it is a commonly used metric, but the problem we highlight here arises from comparing distributions of repeated-measures data without accounting for time, regardless of how the distribution is represented.

Figure 2

Simulation Example Demonstrating the Effect of Varying the Number of Trials in an Experiment



Note. Points show the simulation data. Summarizing the distribution of the data using the typical method of using a normal distribution that disregards time (beige) leads to an estimate that is affected by the early trials, and the size of this effect depends on the number of trials. In early trials, the asymptote estimate (blue) is more uncertain but not shifted toward the early trial like the mean, and as more trials are added, it becomes more certain than the mean, as can be seen in the width of the range of estimates. Full code for simulation is provided in additional online materials at https://osf.io/64r7m/, Section 4. See the online article for the color version of this figure.

another, they tend to vary in the number of replicates, which makes effects sizes smaller or larger, confounding any comparison between them. Rouder and Haaf (2019), in addition to highlighting that the number of trials is at least as important to power and reproducibility as the number of participants, also advocated using hierarchical models that remove variability due to trials, to achieve more portable estimates of effect size. Asymptotic regression also solves the portability problem, but goes a step further by modeling change over the sequence of replicated measurements explicitly, rather than treating each instance as independent. The estimates resulting from hierarchical models would still drift toward the asymptote as the sequence of trials progressed, while the asymptote estimate does not. The examples that follow illustrate the ways in which asymptotic regression can provide an approach to aggregating and describing data that are even more portable than a hierarchical model that ignores time.

Overview of the Examples

In each of the three examples shown below, we demonstrate different aspects of timecourses and how they can be modeled using asymptotic regression, starting with a simple use case and adding experimental complexity with each example. In the first example, asymptotic regression is applied to model change in facial muscle activation taking place over a 5-s interval in response to viewing face change between happy and angry facial expressions (the control condition from Korb et al., 2023). This straightforward case demonstrates how asymptotic regression allows separate estimates of both the rate at which muscles change in response to the face changing expression and the state of the muscles as they reach asymptote.

Example 2 models a well-established tendency to respond more quickly to self-relevant targets, where the association between the participant and an arbitrary shape is introduced at the start of the experiment and then measured over multiple trials. We show that, as the variance and reaction time decrease over time, the differences between conditions emerge more strongly in the asymptotes. This example presents a useful application of asymptotic regression as a way of removing the early effects of practice that can mask the strength of effects. To this end, we show how the estimates of the asymptote would have changed had the experiment included fewer trials, presenting a useful tool for planning efficient experiments that avoid collecting excessive trials of data.

In the final example, we introduce a new set of data that were collected to better understand how visual search behavior changes with experience. Specifically, a well-established measure of visual search performance is the ratio of reaction times on trials where the target is present to reaction times on trials where the target is absent. In this example, we show that the ratio itself decreases to an asymptote over the course of a block of trials, that it resets at the start of each block, and that the rate at which the ratio changes is affected by what the participant experienced in the preceding block. These observations raise a fundamental interpretation question: A typical average ratio, which ignores time, does not accurately reflect behavior, as it only represents a point participants pass through transiently on their way from the start to the end of a block of trials. The ratio therefore depends not only on how many trials the experimenter chooses to include in the experiment, but also on how they split the trials into blocks, and how they intermingle different block and trial types together. This is one example of a ubiquitous problem, considering most effects in the literature are based on averages that ignore time. Asymptotic regression offers a relatively simple solution to these interpretive problems.

Transparency and Openness

All data and analysis code are available on a public Open Science Framework repository at https://osf.io/64r7m/. See the reference list for the full citation (Clarke & Hunt, 2024).

Example 1: Modeling Behavior Within a Trial

For our first example, we demonstrate how asymptotic regression can be used to model the timecourse of behavior changes *within* a trial. The example we are using here comes from Korb et al. (2023), who are interested in facial mimicry. Specifically, they measured facial reactions to dynamic happy and angry faces using the electromyography (EMG) of the zygomaticus major and corrugator supercilii muscles. The original study was interested in how these signals were modulated after pharmacologically manipulating the opioid and dopamine systems, but we will only look at the placebo condition here. A model of the full experiment is presented in the additional online materials at https://osf.io/64r7m/.

Method

A brief summary of the methods is presented below. For further details, please see Korb et al. (2023).

Participants

The full sample included data from 130 participants, although we are only concerned with the subset of 40 who were assigned to the placebo condition. Demographic details of the participants and recruitment are reported in full in an article (Korb et al., 2023). The research was conducted in line with local ethics regulations and the Declaration of Helsinki.

Stimuli

The stimuli consisted of 24 videos based on photos of 10 faces (five male) with happy and angry expressions. The videos were created using morphing software (Morpheus Photo Morpher, Version 3.17) and had a duration of 5 s. The videos displayed a happy face gradually becoming angry, or vice versa, and were repeated four times. This gave a total of 96 trials shown in semirandom order (a maximum of three successive stimuli with the same emotion was allowed) in two blocks of 48 trials.

Procedure

On each trial, participants were instructed to indicate the moment at which the first expression changes into the second, by pressing the arrow-up button on a keypad. A fixation cross was shown at the center of the screen for 2–3 s before each trial, and a feedback screen was displayed after each trial for 1 s. The task was preceded by four practice trials. Details of the EMG measurement procedure can be found in Korb et al. (2023).

Preprocessing

EMG data were preprocessed in Matlab, using the EEGLAB toolbox (Delorme & Makeig, 2004). Epochs were extracted from 0.5 s before to 5 s after stimulus onset, and expressed as a percentage of baseline (the average of the 500 ms preceding stimulus onset). Trials with average or peak values more than 2 *SD*s above or below the mean (for that subject and muscle) were removed from analyses. Trials were also removed if outlier values were detected in the baseline period.

Data were converted to the proportion of baseline (e.g., a value of 0 indicates 0% of the baseline, and a value of 1 indicates 100% of the baseline) and then log-transformed to account for their skewness. The results of two standard approaches are shown in Figure 3: The connected points show the average of the EMG measure within each of five time bins, and the straight line shows a simple linear regression (without applying asymptotic regression).

Analysis

One interesting detail of this study is that each person's EMG signal has been normalized by a baseline signal. This means that we can fix $b = 0$ in our model, leaving us to fit a and r to the data. Both of these parameters are allowed to vary from one condition to the next, and a maximal random effect structure is used (Equation 4):

$$\begin{aligned} a &\sim x_m \times x_e + (x_m \times x_e | Z) \\ b &= 0 \\ r &\sim x_m \times x_e + (x_m \times x_e | Z), \end{aligned} \tag{4}$$

where x_m and x_e are the muscle and emotion variables, and Z is participant ID.

Results

The posterior model fit is illustrated in Figure 4. These are small effects that take time to emerge. At the start of the trial (parameter b), activation is at 0 (baseline). As the participant processes the visual stimulus and their facial muscles react to it, an effect emerges and can be clearly seen in the asymptotes (a). Taking the full set of measurements over the trial would underestimate the effect of facial expressions on these muscles. Comparing asymptotic regression to the standard approaches shown in Figure 3, the change in facial expression over time does not follow a straight line, and in any case, fitting this line does not make interpretation of the data any easier,

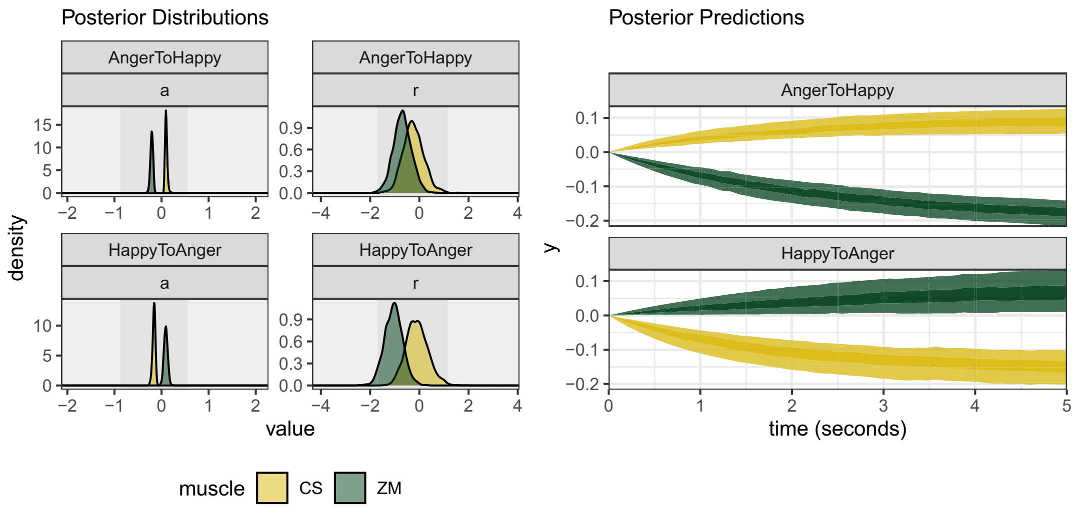
Figure 3

Two Standard Alternative Approaches to Modeling Time: Linear Regression and Binning

Note. Using linear regression like this requires two parameters per condition, so eight in total (not counting random effects). Treating time as a categorical factor would require five parameters per condition (because five bins have been used in this example). It is already clear from the mismatch between the bin averages and the regression line that a straight line is not a good fit with the data. See the online article for the color version of this figure.

Figure 4

Model Fits for the EMG Experiment



Note. All measurements for *a* and *y* are in log units. (Left) These plots show the posterior density estimates for *a* and $r = \text{logit}(\rho)$ for both the angry-to-happy and happy-to-angry conditions. The gray-shaded regions show the 53% and 97% HDPI for the priors. (Right) Posterior predictions. The shaded regions depict the 53% and 97%. EMG = electromyography; CS = corrugator supercilii; ZM = zygomaticus major; HDPI = highest density posterior interval. See the online article for the color version of this figure.

because it does not provide a meaningful end point. An approach that lends itself better to comparison of the conditions is to separate the data into time bins and compare only the later bins, where the effect emerges. But doing so requires making an arbitrary decision about how many bins to use and discards most of the information in the data set. Asymptotic regression makes use of the full time series and provides more sensitive, transparent estimates of the effect of facial expression on muscle activation.

Discussion

The aim of this example is to provide a simple demonstration of asymptotic regression in action. The model presented here is slightly simpler than the one outlined in Korb et al. (2023) as we fix $b = 0$. The other difference is that the original analysis was on the full data set and investigated whether a model that allowed *a*, *b*, and *r* to vary between the three drug groups outperformed a model that did not. The results of this analysis suggested that there was no clear effect of the intervention. Applying asymptotic regression adds confidence that there are not underlying differences obscured by the way data were binned or averaged over time.

Example 2: Self-Reference Effect

In our second example, we fit models to data to illustrate how the self-reference effect (SRE) changes over a block of trials. In the typical SRE experiment (Sui et al., 2012), an association is learned between one's self and a particular object. Participants are then asked to match objects with the labels they learned in a speeded reaction time task. The typical pattern of results is faster RTs to confirm the shape and label match when they were associated with one's own self, relative to the other conditions (when the shape and label do not match, or they matched with the label "stranger" or "friend").

In addition to modeling how the effect changes over repeated trials, we will also demonstrate a more advanced version of the asymptotic regression model in which the variance is allowed to vary over time as well as the mean. See R and Stan code at https://osf.io/64r7m/ for full model specification.

Method

Participants

Twenty-five (16 female, nine male, indicated by selecting from options "M/F/O" in an anonymized form) participants were recruited to a visual search experiment in which the SRE task was included (Bhat et al., 2024). The search aspect of the experiment is not relevant to the current research so will not be described here, except to provide the context that participants completed 102 visual search trials, followed by the SRE task described below, followed by another 102 visual search trials. The full experiment took approximately an hour, and participants were remunerated £10. The protocol was reviewed and approved by the Aberdeen Psychology Ethics Committee.

Stimuli and Apparatus

There were two categories of stimuli: a single line segment tilted 45° clockwise from vertical, which was the object labeled as "YOU" in the experiment, and a set of line segments, each with an orientation that was determined by random selection from a uniform distribution with mean 45° anticlockwise from vertical and a variance of 120. This set of line segments was "STRANGERS." All line segments were 1.2 cm (1.5 visual angle). The display was the same as for the previous experiment.

Procedure

Using a brief on-screen tutorial, participants were taught the associations between the line segment rotated 45° clockwise from vertical and the label "YOU," and sets of four randomly oriented line segments displayed in a single horizontal line and the label "STRANGERS." They were then given 16 practice and 160 experimental trials (note that practice trials were not removed in the asymptotic regression analysis below). On each trial, a single line segment appeared with a label in capital letters underneath ("YOU" or "STRANGER"), and the participant had to indicate whether they matched or not, using the "F" key for match and the "J" key for mismatch. Participants were instructed to respond as quickly and accurately as possible. The four trial types (crossing YOU/STRANGER, Match/Mismatch) were equal in number and randomly ordered.

Preprocessing

Trials with RTs less than 100 ms (12 trials) or longer than 9 s (14) were removed. Two hundred ninety-five incorrect trials were also removed, leaving us with 4,079 observations. In order to produce reliable LOO estimates, a further 26 points (0.6%) were removed.

Analysis

We fit a multilevel asymptotic regression model in which all three parameters are allowed to vary across experimental conditions (Equation 5):

$$\begin{aligned} a &\sim x + (x|Z) \\ b &\sim x + (x|Z) \\ r &\sim x + (x|Z), \end{aligned} \tag{5}$$

where $x \in \{\text{match} - \text{you}, \text{nomatch} - \text{you}, \text{nomatch} - \text{stranger}, \text{match} - \text{stranger}\}$, and Z is participant ID.

Results

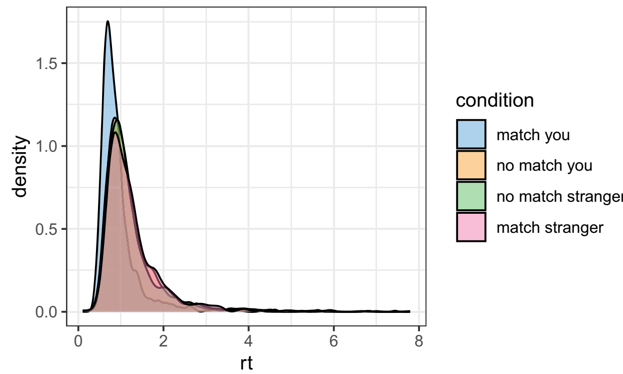
An overview of the RTs is given in Figure 5 while the asymptotic regression model fit to the SRE data is shown in Figure 6. LOO model comparison (see Table 1) suggests that the standard asymptotic model outperforms both linear regression and a stationary normal distribution (7.1%). If we additionally include the varying-variance asymptotic regression model, we see that this is an even better fit of the data.

There is a clear SRE in the asymptotes: The estimated asymptote for the *match–you* condition is clearly lower than the other conditions. We can confirm this result by calculating the posterior probabilities of a difference (see Table 2). The results are more ambiguous as to whether this effect exists in the b and r parameters. This suggests that the SRE only has an effect after participants have spent some time with the task.

We can also see that the variance in log RTs decreases during the experiment. It is worth pointing out that this decrease is above and beyond what we would naturally see when modeling RTs with a lognormal distribution (i.e., a longer μ leads to higher variance when we look at the unlogged data). The empirical data have a standard deviation of 0.45, and the highest density posterior interval estimate for the asymptotic standard deviation is lower, with a range of [0.26, 0.34].

Figure 5

Density Plots Showing the Distribution of Response Times in the Self-Reference Effect Study



Note. See the online article for the color version of this figure.

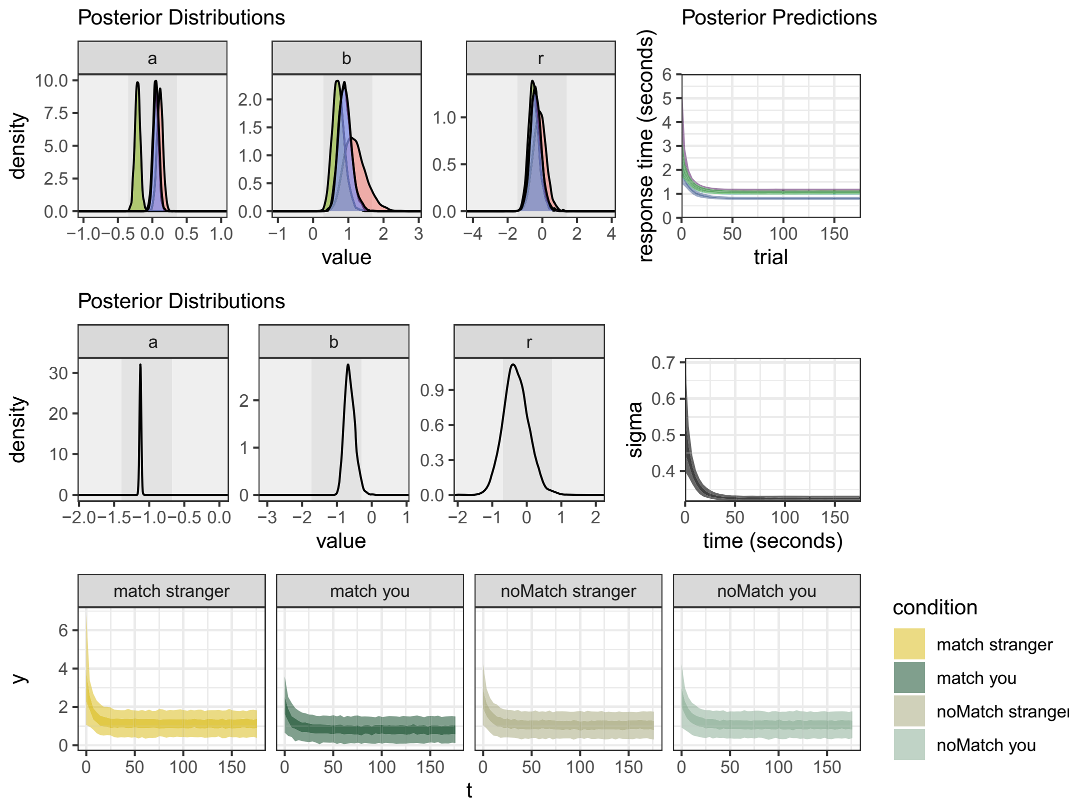
Estimating How Many Trials Are Needed

The fact that the effects emerge only in the asymptotes leads naturally to the question of how many trials we require to accurately estimate the effect. We addressed this by simulating what the model fits would have been if we had collected 25, 50, or 75 trials in total. An important note here is that we are not randomly sampling from the trials, but only fitting the first N trials of the experiment to see how our estimates change with more or less repeated trials in the series. The results are shown in Figure 7, in terms of the distribution of the differences in asymptote between the self and other conditions. We can see the asymptote estimate does not change or become more precise after a hypothetical experiment that contained just 75 trials. This tells us how short the experiment could be to still produce the same results.

Discussion

Two new points are raised by these results. First, the predicted effect of the self-label was relatively weak in the overall reaction times, but stronger in the asymptotes. This implies that the effect emerges after participants learned the task and response mappings and settled into a strategy, demonstrating the potential for using asymptotic regression as a principled way to account for variance in data. These results may also constrain the possible theoretical explanations for the effect since it appears to be a stable behavioral trend that takes some time to emerge but persists over time. Second, the dynamics of the variance can also be modeled using asymptotic regression. This points to a useful application for planning experiments and estimating power. Given an established effect for which data already exist, and knowing that the effect is observed in the asymptotes, it is possible to estimate how power changes with the number of trials to decide how many trials are sufficient to achieve a particular threshold for power.

Figure 6
Model Fits for the SRE Experiment



Note. (Top row) Three plots show the posterior density estimates for a, b, and $r = \text{logit}(\rho)$. The gray-shaded regions show the 53% and 97% HDPI for the priors. The rightmost plot shows how these parameters combine to give the relationship between the trial number and μ. (Middle row) Same as above, but for σ. (Bottom row) Posterior predictions, combining the varying μ and σ. The shaded regions depict the 97% HDPI for the average response time. SRE = self-reference effect; HDPI = highest density posterior interval. See the online article for the color version of this figure.

Example 3: Visual Search

Our final example models new data from a visual search study in which participants were tasked with deciding whether a target was present or not.[6] The aim of this experiment was to better understand a tendency originally observed in Nowakowska et al. (2017) to search for easy-to-find targets for longer than was required. Our working hypothesis was that this tendency to oversearch would diminish with experience and that this would be particularly true when that experience required participants to search under time pressure. This is a hypothesis that asymptotic regression is particularly well-suited to evaluate. Our initial analysis of these data led to our interest in modeling how RTs change over a block of trials and gave rise to the ideas that motivated this article.

The duration of the search display was either unlimited (“long”) or limited to 200 ms (“short”). The question our analysis addresses is: How does RT on easy trials with long durations change as a function of experience? The three key experience conditions are either none (i.e., performance in the first block of trials), experience with long-duration trials (the second block, after completing a full block of “long” trials), or experience with short-duration trials (the second block, after experiencing a full block of short trials). For this example, we demonstrate how we can easily generalize the asymptotic regression model to use nonnormal probability distributions. In this case, rather than fitting a normal distribution to log reaction time, we fit a lognormal distribution to the reaction times directly.

Table 1
LOO Model Weights for the Self-Reference Effect Data for Three- and Four-Way Model Comparisons

Model	Three models	Four models
Stationary μ	0.071	0.087
Linear μ	0.042	0.022
Asymptotic μ	0.887	0.285
Asymptotic μ and σ		0.606

Note. LOO = leave-one-out.

[6] These data were presented at the Scottish Vision Group meeting in April 2023.

Table 2
The Posterior Probability of the Match–You Condition Being Smaller Than Each of the Other Conditions for a, b, and r

Parameter	Match–stranger	No match–you	No match–stranger
a	>0.99	>0.99	>0.99
b	0.97	0.82	0.80
r	0.87	0.67	0.63

Method

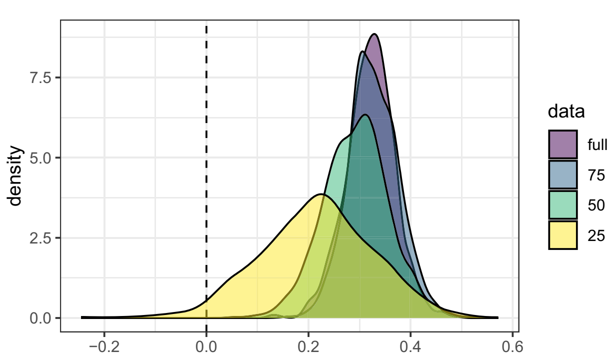
Participants

Fifty people (25 female, 25 male; $M_{\text{age}} = 26$, $SD = 9.3$) completed the experiment. Demographic details were indicated by participants selecting from options "M/F/O" and writing in age on an anonymized form. Full details of the data set and exclusions can be found on https://osf.io/64r7m/. The protocol was reviewed and approved by the Aberdeen Psychology Ethics Committee. None of the participants had taken part in related studies run by our lab.

Stimuli and Apparatus

The stimuli consisted of 22 × 16 arrays of black line segments on a gray background. The target (if present) was the unique item orientated 45° to the right. The remaining line segments had a mean orientation of 45° to the left and a variance of either 30° (easy trials) or 120° (hard trials). For stimulus examples, see Nowakowska et al. (2017). The displays were viewed from a chin rest stationed 50 cm from a 19-in cathode-ray tube ViewSonic Graphics Series G90fB monitor with a 1,024 × 768 resolution and 100 Hz refresh rate. The experiment was programmed using the Psychtoolbox in MATLAB 2014b. Participants were seated alone in a dimly-lit room, and their eye movements were tracked using an EyeLink 1000 desktop-mounted eye tracker (only the keypress RTs are analyzed here).

Figure 7
Estimates for the Asymptotic SRE Effect, Based on Fitting to the First 25, 50, or 75 Trials, or the Full Data Set (176 Trials)



Note. SRE = self-reference effect. See the online article for the color version of this figure.

Procedure

Participants pressed the space bar while fixating a central cross to initiate the eye tracker's drift check and start each trial. The search array for that trial then appeared. The participant was instructed to respond as quickly and accurately as possible by pressing the right arrow key if the target was present or the left arrow key if it was absent. The display remained on until a response. A red screen followed incorrect responses. There were 96 trials in each block, with equal numbers of easy and hard trials and target-present and target-absent trials, randomly intermixed. As the hard trials are essentially filler trials, only data from the *easy* trials are analyzed here.

There were two (blocked) search durations. In *long* blocks, the search display was presented until the participant responded or the trial timed out (maximum of 60 s). In *short* blocks, the search display was presented for 200 ms and then removed, leaving a gray screen until the participant responded or 60 s elapsed. Each participant completed two blocks of 96 trials, in one of three block orders: short then long ($N = 18$), long then short ($N = 18$), long then long ($N = 14$). As noted above, we are interested in modeling how search times change over time. In particular, previous research shows that search strategies are not directly affected by time pressure, but they do show gradual improvements over time (Nowakowska et al., 2021). Our hypothesis was that experience with a block of 200-ms duration trials might improve learning on subsequent long-duration trials relative to equivalent experience with trials with unlimited time. We therefore were particularly interested in comparing the long-duration trials when presented as Block 2, following a block of short-duration trials versus a block of long-duration trials.

Preprocessing

Five participants were removed based on their accuracy and median RT: One participant who failed to achieve at least 75% in the *long* display duration—*target-absent* condition—was removed. We removed a further two participants who had median RTs of over 10 s (the next slowest person had a median RT of less than 5 s), and one person who had a median RT of 3 s for target-present trials (all other participants had a median RT of less than 1.25 s in this condition). This resulted in a total of 43 participants (short then long $N = 16$, long then short $N = 15$, long then long $N = 12$). After these exclusions, accuracy on the easy/long trials was close to 100% so we focus on reaction time in the analysis (see the additional online materials at https://osf.io/64r7m/ for a full description of the data and exclusions).

We also excluded 167 trials (4.1%) with exceedingly long or short RTs (details given in the additional online materials at https://osf.io/64r7m/). Finally, to get reliable LOO estimates, a further 6 points (0.24%) were removed.

Analysis

We fit a multilevel asymptotic regression model in which all three parameters are allowed to vary across experimental conditions (Equation 6):

$$a \sim x_b \times x_{\text{targ}} + (x_{\text{targ}}|Z)$$
$$b \sim x_b \times x_{\text{targ}} + (x_{\text{targ}}|Z) \quad (6)$$
$$r \sim x_b \times x_{\text{targ}} + (x_{\text{targ}}|Z),$$

where $x_b \in \{\text{long first}, \text{long after long}, \text{long after short}\}$; x_{targ} represents whether the target is present or absent; and Z is participant ID.

Results

As with previous examples, the LOO model weights strongly favor the lognormal asymptotic regression model (84.6%) over linear (9.1%) and stationary (6.3%) models. From the posterior predictions, for target-present trials, the search task is indeed easy, with the average participant taking around 1 s to respond, and there is little evidence of any change in performance over the block. The target-absent trials, in contrast, change substantially over time. For visual clarity, we have therefore focused on target-absent RTs in Figure 8. From this figure, we can see that (a) target-absent RT asymptotes are similar across conditions, but somewhat faster in the second block after experiencing short durations in the first block. (b) The initial value is higher for the first block compared to the second, reflecting the effects of practice with the task. There is also a notable "reset" effect in the initial values at the start of Block 2. This is clearest in the posterior predictions on the bottom half of Figure 8: The shift from Trial 96 to 97 represents the end of one block and the start of the next, and there is an uptick followed by a return toward asymptote. (c) There is an intriguing difference in the rate at which performance improves over trials as a function of experience. In particular, practice with long-duration trials in the first block is associated with a faster drop toward asymptote in the second block compared to when there is a change from short to long durations.

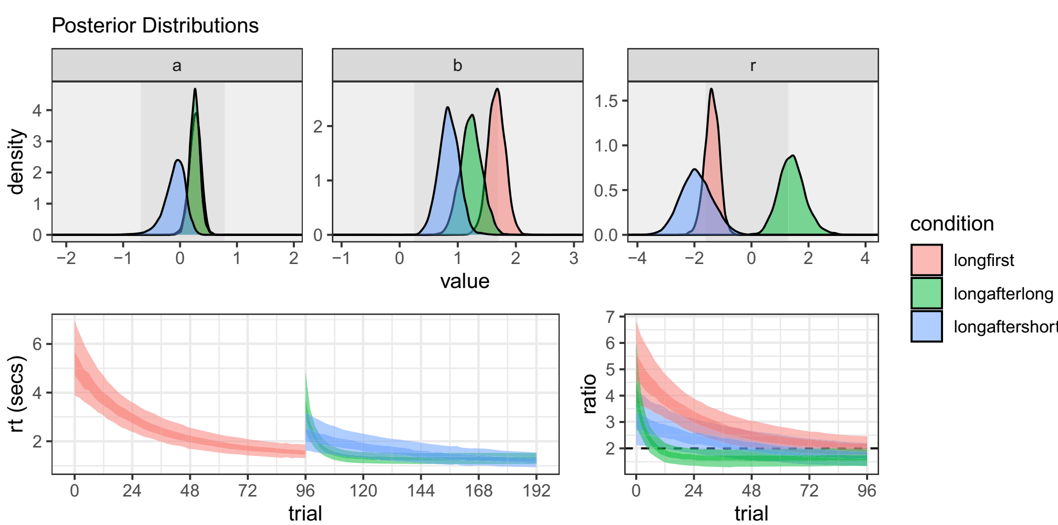
A particularly important metric of visual search performance is the ratio of target-absent to target-present RTs. It has long been noted in the literature (e.g., Treisman & Gelade, 1980) that we can expect ratios to be around 2:1, if participants carry out an exhaustive item-by-item search. This is because for target-absent trials, each item will be inspected once, and for target-present trials, the target will be found on average after inspecting half of the items. But it has also been observed that these ratios are nearly always larger than this in empirical data (Wolfe, 1998). The bottom right panel of Figure 8 provides a tentative explanation by illustrating how the ratio of target-absent to target-present reaction time changes over the course of the block of trials, with 2:1 denoted as the horizontal dashed line. The ratio approaches 2:1 in the asymptotes, but because of the gradual learning effects in the target-absent trials, early in the block the ratio is much higher than this. If instead we had averaged the early and later trials, as is done in the typical approach to calculating this ratio, we would have concluded overall that the ratio exceeds 2:1. Moreover, the amount by which our estimates would exceed 2:1 would depend on how many trials we decided to run, how these trials were blocked, and the order in which blocks were presented. This reinforces, in real data, the problem raised in the simulations presented in the introduction: In the presence of practice effects, estimates of performance that disregard the trial sequence are problematic for comparing across conditions and studies that might differ in seemingly small methodological details like how many practice trials were included. In contrast, the asymptotes provide estimates that are stable and more directly comparable from one experiment or condition to another.

Discussion

In this example, we applied asymptotic regression to uncover the effects of experience on visual search times. Using this method, we could more precisely describe how target-present and target-absent RTs, and the ratio between them, changed from the start of each block to the end. The separate measurement of initial value, rate of change, and asymptote allowed us to observe the specific effects of experience with different conditions on performance, with implications for

Figure 8

Asymptotic Regression Fit for the Search Duration Data



Note. (Top row) Posterior distributions for a, b, and $r = \text{logit}(\rho)$ for the visual search data, target-absent data only. Note that a and b are in log seconds. (Bottom left) Posterior predictions for the mean reaction time. Shaded regions indicate 53% and 97% HDPI. (Bottom right) Posterior predictions for how the target-absent versus target-present (TA:TP) response time ratio changes over time. HDPI = highest density posterior interval. See the online article for the color version of this figure.

understanding the source of search inefficiencies. Had we instead applied standard analyses to these data, we would not have been able to distinguish or estimate the distinct effect of experience on the rate of performance changes over time.

This example also underscores the cautionary point we make in the introduction, that effects that change over time cannot be expressed as a single number. Without taking the trial sequence into account, we would simply conclude that responses are slower on target-absent trials than a 2:1 ratio with target-present trials would predict. After considering RTs in an appropriate temporal context, however, we can see that this conclusion does not describe the data very precisely or completely. The way the ratio between target-present and target-absent trials changes over trials is particularly striking because, as a ratio, this measure removes the overall practice effect on RT and presents the relative change between two conditions. An important implication of these results is that if we had summarized our data using the typical method of averaging all trials together, the amount by which the 2:1 ratio is exceeded would have depended critically on how many trials we had included in the original experiment, and on how many practice trials we ran. Based on fewer trials and less practice, the ratio could have been 3:1 or 4:1. Based on more trials and practice, it would look like 2:1, or possibly even lower. Neither conclusion would have been correct, given that they both ignore the substantial effect of the repeated sequence of trials.

Limitations

In this final section, we outline four potential limitations of using asymptotic regression, along with proposed solutions. While these limitations are all discussed in the context of asymptotic regression, it is worth bearing in mind that they are comparable to the limitations associated with applying any of the more standard statistical models inappropriately.

No Change Over Time

Sometimes there is no trend in the data. This can happen frequently with real-world data: Perhaps the task is very easy or hard, and participants either start at ceiling performance or fail to improve from some floor. This can also occur if participants are already at their asymptotic performance at the onset of data collection. For example, it is common practice to include a number of practice trials when collecting data from participants (human and other animals). In these situations, if practice data are discarded, the learning has already occurred before we start with Trial 1, and the asymptote may already have been reached.

Such situations cause problems for our modeling procedures, similar to issues around collinearity when dealing with linear models. Ideally, we should conclude that $a = b$, that is, the initial value is already at asymptote. However, this leads Equation 2 to simplify:

$$\mu = a - (a - b)(1 - \rho)^{t/t_s} = a. \tag{7}$$

This leaves ρ unconstrained, which can cause convergence problems for model fitting procedures. The problem is actually somewhat worse than this, as we have no guarantee that we will conclude that $a = b$ from the data. For example, as $\rho \to 1$ (i.e., as estimates of ρ approach 1), $(1 - \rho)^{t/t_s} \to 0$. This leads Equation 7 to reduce to $\mu = a$, leaving us with b unspecified. Using similar logic, as $\rho \to 0$, $(1 - \rho)^{t/t_s} \to 1 \Rightarrow \mu = b$ leaving a unspecified.[7] In summary, when the data are constant, we estimate either $a = b$ with ρ unspecified, $a, \rho \approx 1$ with b unspecified, or $b, \rho \approx 0$ with a unspecified.

One benefit of using a Bayesian framework for statistical inference is that we can elegantly tackle this problem by selecting an appropriate prior for $r = \text{logit}(\rho)$ that places the bulk of prior probability away from the problematic floor and ceiling values. An example of this is given in Figure 9 in which we have simulated data with $a = b = 1$. We fit two models to the data, one using a prior of $r \sim N(0, 1)$ while the other uses $r \sim N(0, 5)$. The $N(0, 1)$ gives (after applying the logit transform) a 95% interval of [0.125, 0.875]. This contrasts with the $N(0, 5)$ prior which leads to a *U*-shaped distribution with most of the weight bunched up around 0 and 1. The 97% highest density posterior intervals for the posterior distributions are given in Table 3, and we can clearly see that using a wider prior for ρ leads to increased uncertainty in the estimates for a and b and a bimodal posterior estimate for ρ. A similar solution could be applied under a frequentist framework, by setting boundaries on possible values of ρ.

Linear Trend

While a linear trend is unlikely over a large enough collection of trials (presumably there is some limiting factor on how high or low response variables can be, whether due to physical or biological constraints), it is quite possible that a trend appears to be linear in the observed/collected data for low numbers of trials. An example of an asymptotic fit to linear data is given in Figure 10. While we end up with a reasonable fit over the range of observed data, the model diverges from the linear groundtruth behavior as t increases. Furthermore, we can see from Figure 10 that there are strong correlations between the posterior samples of some of our model parameters. Such correlations are a classic indication that care should be taken with the parameter estimates as there are multiple ways in which we can adjust a and r to give similar predictions.

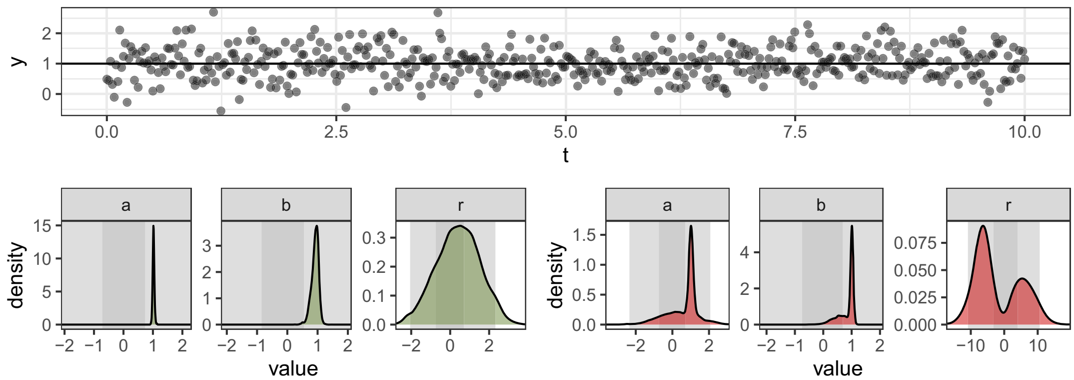
If you find yourself in such a situation, then we suggest either collecting data over a longer period of time or using an informative prior for a (or r). Alternatively, you could simply use linear regression if you believe that the process you are studying is truly linear.

Nonmonotonic Trend

The previous two examples were cases in which we were fitting our three-parameter model to data that were generated by either one (constant) or two (linear) parameter models. While this can lead to problems around overfitting and correlated parameters, both these problems can largely be addressed by selecting suitable priors. An alternative set of problems can arise from underfitting, where we might struggle to fit our monotonic asymptotic regression model to data that have a more complex structure. For example, a participant's performance may initially improve as they familiarize themselves with the task but then decline due to fatigue or boredom. Another possibility is that if we split our trials up into a number of blocks, we

[7] In the first case, the rate is so high that we reach the asymptote by the time we get to the first trial. In the second, the rate is so low that we do not measurably move away from the initial value by the time we get to the final trial.

Figure 9
Example Model Fit to Simulated Data With No Trend



Note. (Top) Data generated from a $N(1, 0.5)$ distribution. (Bottom left) Model fits with a $N(0, 1)$ prior for $r = \text{logit}(\rho)$. (Bottom right) Model fits with a $N(0, 5)$ prior for r. We can clearly see that the $N(0, 5)$ leads to the posterior estimates of r to be bimodal, leading to problems in estimating the other parameters. See the online article for the color version of this figure.

may see steady improvements within each block, but with a partial "reset" to a new initial starting point at the start of each new block (as observed in Example 3, although we fit blocks separately in this example).

These more complex trends present problems for our method (as well as for commonly used methods such as analysis of variance or multilevel linear models). In such cases, more complex time series models could allow for nonmonotonic behavior, or one could use an atheoretical approach such as fitting splines.

Variation in Time Range and Frequency

In the three examples we presented here, the time range and frequency were the same for every participant. The first example sampled muscle activation at a fixed rate over a time interval that was the same on every trial. In the second and third examples, the time unit was trial, and every participant started at Trial 1. But in many other instances, the sampling range and rate over time can vary. For example, over the span of a single trial, the influence of different stimulus characteristics waxes and wanes (van Zoest et al., 2010), and so it can be critically important to consider the time at which responses were executed when trying to understand or compare conditions or response modalities that have different latencies (e.g., Hunt et al., 2007). To show how behavioral biases and errors change with RT, the typical approach is to allow for spontaneous variation in RTs in a group of participants performing a series of trials, and then to treat RT as an independent variable. In the additional online materials at https://osf.io/64r7m/, we provide an example of this approach from an eye movement control experiment, in which we model changes in the bias to fixate a salient distractor (known as oculomotor capture; Theeuwes et al., 1998) over time. The limitation here comes from the issue that each participant produces a different time series, meaning there is a unique "Time 0," time scale, and distribution of samples over that timescale for each individual. We can apply asymptotic regression to individual binary data and estimate the probability of fixating the target as a function of when the saccade was initiated for each person separately, and this can provide a useful set of parameters at the level of the participant, which could be analyzed just as any other participant-level summary statistics. But applying a multilevel version of this analysis is not straightforward. There are possible statistical solutions to this problem (such as rescaling time or implementing some strong priors) but they come with tenuous assumptions. To be confident in the results, we would require a more strictly time-controlled experiment and/or a much larger data set. To be fair, however, this is a potential limitation of any multilevel model where there is a lack of control over how participant data are distributed across the conditions.

Table 3
97% Highest-Density Posterior Intervals for Model Fits to Data With No Trend

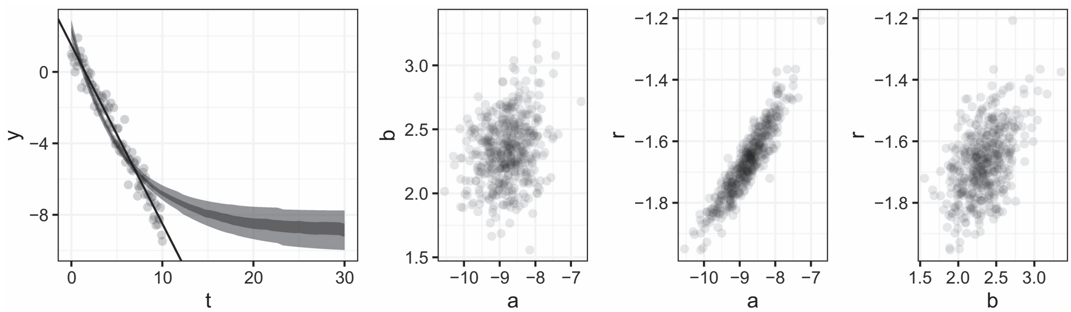
σ	*a*	*b*	*r*
1	[0.9, 1.08]	[0.79, 1.33]	[−2.43, 2.68]
5	[−1.46, 2.50]	[0.139, 1.11]	[−13, 11.8]

General Discussion

Changes in human behavior over time can often be described as a monotonic increase or decrease toward an asymptote. We presented three illustrative examples that demonstrate how asymptotic regression can be applied to estimate the start point, rate of change, and asymptote of behavior that takes this form. Our goal is to encourage broader use of this statistical approach, as it provides a useful and versatile convention for describing the temporal profile of many behaviors.

In the first example, we modeled changes over a timescale of a few seconds. Modeling the changes in EMG over the course of a single trial revealed the subtle emergence of the effects of viewing changes in facial expressions in the asymptote estimates that provided a more precise baseline to which to compare the effects of pharmaceutical interventions. In the second and third examples, participants completed blocks of repeated trials, as is common practice in behavioral experiments. Asymptotic regression allowed us to estimate how the effects of interest changed from the start of

Figure 10
Example Model Fit to Simulated Data With a Linear Trend



Note. (Left) Data generated from a $N(\mu, 0.5)$ distribution, where $\mu = 1.5 - t$. The shaded region shows the fit of the asymptotic regression model, while the line shows the linear groundtruth. (Other plots) Pairs plots of samples from Markov chain Monte Carlo sampling. We can see the considerable correlation between a and r.

the block of trials to the end. In one case (the SRE), the differences in RTs between self-relevant and control objects strengthened over repeated trials. As a consequence, the small, noisy effect in the overall averages was more clearly evident in the asymptotes, demonstrating the potential of asymptotic regression as an approach to improving signal-to-noise ratios in repeated-measures data, even when time effects are not of direct theoretical interest. RT differences that reflect a stable behavioral effect can be masked by the larger variance observed early in the experiment. We use this observation to show how the estimates of the strength of the self-relevance effects would have changed as the number of trials increased, and propose using asymptotic regression for planning experiments that are maximally sensitive and efficient.

In the third example, we use asymptotic regression to describe how visual search performance changes with experience. Separately estimating the initial value, rate of change, and asymptote parameters allows for a more accurate description of behavioral dynamics than conventional approaches. An important observation to highlight from this example was the substantial change in estimates of the ratio of RTs on target-absent versus target-present (TA:TP) trials over the sequence of trials. This is important because it demonstrates that estimates of the size of the TA:TP ratio based on medians or means would depend on how many trials are run: The ratio would get smaller as the number of trials increased. Conclusions about the mechanisms of visual search are often grounded in whether the TA:TP ratio is larger or smaller than 2:1, and how it deviates from this ratio across different conditions (e.g., Wolfe, 1998). But which estimate is the "true" ratio of target-present to target-absent reaction time in these example data? Is it the average ratio, the ratio at asymptote, or the ratio at the start? To consider any of these in isolation will misrepresent the data, and so we advocate for the full function being the estimate, as opposed to any one parameter in isolation.

Asymptotic regression applied to a series of trials can perform a range of functions that are both theoretical and practical. On the theory side, the timecourse of effects can help us understand their fundamental nature. Do they reside largely in the way people respond to a new or unfamiliar context, or do they reside in how people respond after they have adapted to that context? When generalizing an effect from one context to another, for example, from a laboratory context to a more ecologically relevant context, we should expect a phenomenon that only appears or stabilizes after many repeated trials (as was the case for the SRE in our second example) to have less potent effects in everyday situations than one which is evident at Time 0. At the same time, effects that abide in the asymptotes suggest a stable behavioral characteristic, and as such, the extent to which different conditions can change asymptotic effect sizes may be more useful and convincing evidence for theory development relative to changes in overall means or in initial values.

On the more practical side, asymptotic regression provides a principled and reproducible way to boost signal and reduce noise in data, particularly for effects that depend on learning a complex set of responses, acquiring a set of expectations, or adjusting task sets or attentional control settings, which may take some time to instantiate. This may be especially useful for comparing groups of participants who may be more or less experienced in the testing conditions (e.g., comparing across age cohorts or neurodiverse samples), as these differences will reside in the start points. We also demonstrated that asymptotic regression can be useful for planning experiments, by guiding our choices about how many trials we need to run before our effects can be expected to reach asymptote. Just as we can decide how many participants to sample using power analysis on preexisting data, we can apply asymptotic regression to avoid wasting precious participant and equipment time by running more (or less) trials than we need to get a stable estimate of performance levels that approach an asymptote.

A related persistent challenge in behavioral research is the problem of whether and how to exclude outliers in data. The motivation for removing extreme values is to improve the signal-to-noise ratio in detecting true differences between conditions or groups, but two recent articles in this journal (Karch, 2023; Miller, 2023) have made compelling cases against outlier exclusions. In the case of between-group comparisons, Karch (2023) argued that removing outliers leads to inflated effects sizes, an increased risk of Type I errors, and uninterpretable confidence intervals. For reaction time data specifically, Miller (2023) compared 58 outlier exclusion procedures and concluded they do not, on balance, improve power and run the risk of biasing results. In both cases, the recommendation is only to remove data we can be confident are "erroneous." We note that asymptotic regression does not offer a solution to the problem of outliers. But it does improve signal-to-noise ratios, by attributing a portion of what would otherwise be considered "noise" to the effects of time, leading to more precise estimates of effect sizes. We generally agree that only the most extreme cases should be removed from data, and this is the approach we adopted in the analyses reported here.

In the examples we present here, we model the effect of time on continuous measures, with a single measure from each timepoint of the series (e.g., measuring the effect of repeated trials across a block, where there is one reaction time per trial). A related challenge here comes from applying asymptotic regression to parameters derived from distributions of data; for example, probability in the example in the additional online materials at https://osf.io/64r7m/, or d-prime (Cochrane & Green, 2021) and drift-diffusion model outputs (Cochrane, Sims, et al., 2023). The difficulty comes from fitting complex models over multiple trials: Each person only has one first, second, third (and so on) trial, and these have to be shared out between all the conditions in the experiment. It is theoretically possible with a good set of strong, informed priors and an adequate sample size, but these are potential limiting factors on the hierarchical application of asymptotic regression to parameters whose estimation relies on the full set of trials. For examples of solutions to these challenges, the articles referenced above provide an excellent starting point.

Another limiting factor is that asymptotic regression does not, of course, fit all behavioral data. We illustrated the problems that can arise from trying to fit asymptotic regression to data that are flat or linear over time. In both of these cases, using means/medians or linear regression, respectively, may be more appropriate, but it is also instructive to see what happens when asymptotic regression is applied instead. This may sometimes be necessary, as illustrated in our third example set of data, in which we had one condition (target-absent) in which there was a clear downward trend as reaction times got faster over time and approached an asymptote, and another (target-present) which was already near the floor at the start of the experiment, and so remained relatively flat over time. The model estimated a start point and asymptote of both target-present and target-absent conditions, but was highly uncertain about the rate for the target-present condition in particular. The behavior of the model in this case was reasonable, and we were still able to compare the two conditions, but as we cautioned in the Limitations section, problems could arise in the other two parameters if priors around the rate are too wide. More generally, to accurately estimate the full timecourse, it is crucial to have measured it in the first place. Cutting short the data collection interval before performance levels off may result in a strictly linear trend, and the asymptote can no longer be reliably estimated. Similarly, failing to record or include practice data will result in unrealistically flat performance over time. This will make estimates of the start point and rate meaningless. In both cases, crucial information about the data will be missing from the analysis.

There are many other ways in which time can have systematic effects on data that require different approaches from asymptotic regression. For example, performance can build up or break down as conditions repeat or switch, such as in repetition priming and repetition suppression (e.g., Bertelson, 1961; Wiggs & Martin, 1998), and these are best approached by coding trials according to what preceded them. Oscillations at a variety of periodicities can be measurable in an experimental context, such as α waves (8–12 Hz), or basic-rest activity cycles (around 90 min). Here, spline-fitting is the more appropriate tool. Spline-fitting could be considered a more general-purpose tool than asymptotic regression because it can describe performance changes that fluctuate rather than following a monotone. However, the parameters from fitting splines are not as easily interpreted, nor as comparable across different experiments, so we think routinely using asymptotic regression as a general tool for describing a wide range of different behavioral data is a more promising approach.

Many questions related to the process of designing an experiment to measure behavioral effects do not have clear principles to guide the answers. How many practice trials is enough? How many experimental trials should be run? Asymptotic regression can provide much-needed consensus for making these decisions, as demonstrated in these examples. There need be no distinction between practice and experimental trials if the full set of data contributes to the analysis, and as demonstrated in Example 2, previous data can tell us how many trials we can expect will be needed to achieve a stable estimate of a given effect. Being able to assign some of the variability in data to a temporal source also means less unexplained variance in the denominator of our effect size calculations. Just as we routinely account for variability due to participants in our statistical models, we could also be routinely accounting for variability due to time and factoring it out of our behavioral effects when it is not of direct interest. Asymptotic regression is just one of many possible models for describing time effects, and a more general point is that any approach to accounting for the temporal characteristics of data is better than none. Our examples demonstrate that ignoring timecourse effects in data, where they exist, can lead to an impoverished, and sometimes even inaccurate, view of behavior. Our goal in the current article is to raise awareness of asymptotic regression as a relatively simple and versatile tool for describing a common profile of timecourse effects: a monotonic change toward an asymptote. This tool has a long history but has not become part of the common statistical arsenal used in analyzing behavioral data. Having a shared approach and language for estimating and describing this common temporal profile would be an important first step toward a more complete and accurate view of behavior across all areas of psychology.

Constraints on Generality

Asymptotic regression can be applied to any data that take the form of a monotonic increase or decrease toward a stable asymptote. As noted in the Limitations section, there are many other forms the data can take, for which other approaches are more appropriate.

References

Bertelson, P. (1961). Sequential redundancy and speed in a serial two-choice responding task. *Quarterly Journal of Experimental Psychology*, *13*(2), 90–102. https://doi.org/10.1080/17470216108416478

Bhat, M., Nowakowska, A., Clarke, A. D. F., & Hunt, A. R. (2024). You are not enough: Inefficient search strategies persist for self-relevant targets[1]. *Visual Cognition*, *32*(2), 135–150. https://doi.org/10.1080/13506285.2024.2393890

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71. https://doi.org/10.1006/cogp.1998.0681

Clarke, A. D., & Hunt, A. R. (2024). *Learn more from your data with asymptotic regression: Data, analysis scripts and supplementary information*. https://osf.io/64r7m/

Cochrane, A., Austerweil, J. L., Simmering, V. R., & Green, C. S. (2018). Rapid learning in early attentional processing: Bayesian estimation of trial-by-trial updating. In C. W. Kalish, M. Rau, J. Zhu, & T. T. Rogers (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 232–237). Cognitive Science Society.

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

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

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

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*(1), 585–612. https://doi.org/10.1146/annurev.psych.49.1.585

Heathcote, A., & Brown, S. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, *7*(2), 185–207. https://doi.org/10.3758/BF03212979

Hughes, A. E., Nowakowska, A., & Clarke, A. D. F. (2024). Bayesian multilevel modelling for predicting single and double feature visual search. *Cortex*, *171*, 178–193. https://doi.org/10.1016/j.cortex.2023.10.014

Hunt, A. R., von Mühlenen, A., & Kingstone, A. (2007). The time course of attentional and oculomotor capture reveals a common cause. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(2), 271–284. https://doi.org/10.1037/0096-1523.33.2.271

Karch, J. D. (2023). Outliers may not be automatically removed. *Journal of Experimental Psychology: General*, *152*(6), 1735–1753. https://doi.org/10.1037/xge0001357

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), Article 3. https://doi.org/10.1167/17.11.3

Kershner, A. M., & Hollingworth, A. (2023). Category-specific learning of color, orientation, and position regularities guide visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *49*(6), 907–922. https://doi.org/10.1037/xhp0001098

Koch, I. (2003). The role of external cues for endogenous advance reconfiguration in task switching. *Psychonomic Bulletin & Review*, *10*, 488–492. https://doi.org/10.3758/BF03196511

Korb, S., Clarke, A., Massaccesi, C., Willeit, M., & Silani, G. (2023). Facial mimicry is not modulated by dopamine D2/3 and opioid receptor antagonism. *Psychopharmacology*, *240*, 2081–2091. https://doi.org/10.1007/s00213-023-06426-3

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001. https://doi.org/10.1016/j.jmva.2009.04.008

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492–527. https://doi.org/10.1037/0033-295X.95.4.492

Manassi, M., Murai, Y., & Whitney, D. (2023). Serial dependence in visual perception: A meta-analysis and review. *Journal of Vision*, *23*(8), Article 18. https://doi.org/10.1167/jov.23.8.18

Miller, J. (1988). A warning about median reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 539–543. https://doi.org/10.1037/0096-1523.14.3.539

Miller, J. (2023). Outlier exclusion procedures for reaction time analysis: The cures are generally worse than the disease. *Journal of Experimental Psychology: General*, *152*(11), 3189–3217. https://doi.org/10.1037/xge0001450

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

Nowakowska, A., Clarke, A. D. F., & Hunt, A. R. (2017). Human visual search behaviour is far from ideal. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1849), Article 20162767. https://doi.org/10.1098/rspb.2016.2767

Nowakowska, A., Clarke, A. D. F., von Seth, J., & Hunt, A. R. (2021). Search strategies improve with practice, but not with time pressure or financial incentives. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(7), 1009–1021. https://doi.org/10.1037/xhp0000912

Pasqualotto, A., Cochrane, A., Bavelier, D., & Altarelli, I. (2024). A novel task and methods to evaluate inter-individual variation in audio-visual associative learning. *Cognition*, *242*, Article 105658. https://doi.org/10.1016/j.cognition.2023.105658

Rast, P. (2011). Verbal knowledge, working memory, and processing speed as predictors of verbal learning in older adults. *Developmental Psychology*, *47*(5), 1490–1498. https://doi.org/10.1037/a0023422

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, *26*(2), 452–467. https://doi.org/10.3758/s13423-018-1558-y

Speekenbrink, M., & Shanks, D. R. (2010). Learning in a changing environment. *Journal of Experimental Psychology: General*, *139*(2), 266–298. https://doi.org/10.1037/a0018620

Stan Development Team. (2020). *RStan: The R interface to Stan* (Version 2.21.2) [Computer software]. https://mc-stan.org/

Stevens, W. L. (1951). Asymptotic regression. *Biometrics*, *7*(3), 247–267. https://doi.org/10.2307/3001809

Sui, J., He, X., & Humphreys, G. W. (2012). Perceptual effects of social salience: Evidence from self-prioritization effects on perceptual matching. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(5), 1105–1117. https://doi.org/10.1037/a0029792

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science*, *9*(5), 379–385. https://doi.org/10.1111/1467-9280.00071

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

van Zoest, W., Hunt, A. R., & Kingstone, A. (2010). Representations in visual cognition: It's about time. *Current Directions in Psychological Science*, *19*(2), 116–120. https://doi.org/10.1177/0963721410363895

Vehtari, A., Simpson, D., Gelman, A., Yao, Y., & Gabry, J. (2024). Pareto smoothed importance sampling. *Journal of Machine Learning Research*, *25*, 1–58. https://jmlr.org/papers/v25/19-556.html

Wang, B., & Theeuwes, J. (2018). How to inhibit a distractor location? Statistical learning versus active, top-down suppression. *Attention,* *Perception, & Psychophysics*, *80*, 860–870. https://doi.org/10.3758/s13414-018-1493-z

Whitehead, A. N. (1929). *Process and reality*. Macmillan.

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, *8*(2), 227–233. https://doi.org/10.1016/S0959-4388(98)80144-X

Williams, D. R., Zimprich, D. R., & Rast, P. (2019). A Bayesian nonlinear mixed-effects location scale model for learning. *Behavior Research Methods*, *51*(5), 1968–1986. https://doi.org/10.3758/s13428-019-01255-9

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, *9*(1), 33–39. https://doi.org/10.1111/1467-9280.00006

Received October 20, 2023
Revision received October 17, 2024
Accepted October 24, 2024 ■