# Longitudinal Computational Modeling

## 1 Introduction

Across multiple assessment strategies(e.g., self-reports, decision-making tasks), conventional scoring procedures frequently characterize patterns of behavior by overall aggregate indices. However, a critical challenge for these measures is that individual behavioral outputs are produced via multiple competing and complementary psychological processes (Wiecki et al., 2015). While conventional scoring procedures are unable to discriminate between these processes, more recently developed generative models of behavior are well-positioned to discriminate between individual psychological processes, yielding enhanced specificity in behavioral metrics as well as improved tasks psychometrics and better clarity for probing associations with psychological processes (Ahn et al., 2017; Chen et al., 2015; Huys et al., 2016). The use of generative modeling is becoming commonplace in multiple fields of study, with the large majority of studies relying on cross-sectional designs. There is a growing number of studies that have begun examining test-retest reliability across a small number of assessment waves (Karvelis et al., 2023). However, for applications of generative modeling to understand longer-term change across multiple contexts, including naturalistic course or in the context of intervention, more than two assessment occasions are needed. Longitudinal changes are frequently examined using growth models, often estimated using linear mixed models. Thus far, generative modeling frameworks and longitudinal multilevel models (Bryk & Raudenbush, 1992) have yet to be integrated. Here, we illustrate a means of estimating longitudinal changes in parameters from generative models in a single modeling framework. We demonstrate this model estimation first from a simulation of a single parameter reward learning model and then extend this demonstration to a real-world example of longitudinal changes in the Iowa Gambling Task across a five-wave longitudinal study.

### Longitudinal Modeling

Longitudinal research takes many forms with a major focus on test-retest reliability. Typically, studies examining test-retest reliability assess mean-level and rank-order consistency of task performance using traditional scoring procedures (e.g., Buelow & Barnhart, 2018; Schmitz et al., 2020). Studies examining reliability for generative model parameters are less common but are becoming an increasing area of focus given concerns regarding the reliability of these parameters (Karvelis et al., 2023). Recently, joint modeling, in which model parameters and their reliabilities across multiple assessments are estimated simultaneously, offers a means to improve reliability with multiple studies supporting this approach for two timepoint data (Haines et al., 2020; Haynes et al., 2024; Sullivan-Toole et al., 2022). We discuss this approach in further detail later. In the context of assessing test-retest reliability, however, the only means of evaluating change involves a simple difference or change from one occasion to another (Ployhart & MacKenzie, 2014).

In the context of longitudinal changes in development, course of treatment, or intervention outcomes, studies frequently employ more than two assessments, which provides flexibility in the modeling of change across time, including changes in mean-level and rank-order stability in the same model. Some methods, such as repeated measures analysis of variance (RM-ANOVA), offer a means of testing mean-level differences between assessments. RM-ANOVA is typically implemented by estimating simple mean-level differences and is unable to accommodate missing data without the use of other methods (e.g., multiple imputation). Other methods, including multilevel models (MLMs) and latent growth curve models (LGCMs), provide additional flexibility for considering underlying trajectories of change that explain the mean-level changes in outcomes. Despite their differences in data organizational structures, the estimation of MLMs and LGCMs are identical when requisite constraints are applied. The trajectories are characterized by point estimates of starting points (i.e., intercepts) and rates of change (i.e., slopes), as well as random effects reflecting individual differences in intercepts and slopes. Studies examining longitudinal changes in performance on behavioral tasks frequently use summary measures calculated from traditional scoring procedures (e.g., choice proportions) for these models. As noted above, however, these indices may conflate the multiple processes leading to the behaviors.

At its core, generative modeling involves fitting a theoretical model of behavior, and such models have been used in longitudinal studies to assess change across time in task performance. For example, Klein et al. (2022) used the hyperbolic discounting model to estimate the degree of discounting, a measure of impulsive decision-making, in a sample ranging from childhood and adulthood. Estimates of discounting were then used to examine longitudinal changes in impulsive decision-making across development, and they found that the degree of discounting decreases rapidly in early childhood and then stabilizes in mid-to-late adolescence. Two-stage approaches such as this are frequently used when we use theoretical models of behavior to make inferences about a population via a statistical model, and they have proved useful in understanding longitudinal change, illustrated by Klein et al. A disadvantage of two-stage approaches, however, is that estimates from the theoretical model are treated as “true” scores (i.e., observed without error) in the statistical model, and this can bias inference (**need citation here**). Thus, an approach that incorporates the uncertainty associated with estimating theoretical parameters into the statistical model (e.g., longitudinal MLM) could improve our overall ability to estimate longitudinal changes in computationally-derived parameters.

Researchers examining longitudinal changes in these behavioral processes do so in two-stage approaches. Specifically, a behavioral model is fit to the data at each timepoint separately, and then a longitudinal model is fit to the parameters from the behavioral model. Such an approach has yielded important insights regarding how some behavioral processes develop across time.

Here, we illustrate how to incorporate the uncertainty associated with estimating theoretical parameters within a statistical model via hierarchical Bayesian modeling. Hierarchical modeling allows us to use information derived from all participants and all timepoints to inform estimates of different individuals and at different timepoints. Bayesian estimation is a more flexible and more powerful approach to hierarchical modeling that is well-suited for estimating computational and growth curve parameters.

## 2 Simple Longitudinal RL Model

To illustrate the longitudinal computational modeling framework, we begin with a simulated example of how to construct such a model. We first constructed a hypothetical two-armed bandit task, a task characterized by two options that participants can choose from across a series of trials, with each choice resulting in either a reward (+1) or no reward (0). Here, the probability of reward was 30% and 70% for options 1 and 2, respectively, and there were 100 total trials. Next, we built a one-parameter longitudinal reinforcement learning model to simulate data for the task across four conditions. Choices within the task are assumed to be drawn from a Bernoulli distribution, such that

|  |  |  |
| --- | --- | --- |
|  |  | Equation 1 |

where *Yi,s*(*t*) is the choice for either option 1 (*Y* = 0) and option 2 (*Y* = 1) on trial *t* by participant *i* on session *s*, and *V1,i,s*(*t*) and *V2,i,s*(*t*) are the expected values associated with choosing option 1 or 2, updated from trial to trial according to the following function:

|  |  |  |
| --- | --- | --- |
|  |  | Equation 2 |

where *A* is a free parameter describing learning rate for both options, and *x*(*t*) is the amount of the outcome on trial *t*. Learning rates are bounded between 0 and 1. Equations 1 and 2 represent a simple reinforcement learning model describing how rewards affect choices for the two options.

For the longitudinal computational model, we estimated the learning rate using a growth function. To do this, we estimated an unbounded learning rate, `*A*, that was subsequently transformed such that *A* = Φ( `*A*), where Φ is the cumulative distribution function of the standard normal distribution. This transformation ensures that the learning rate used in the computational model is between 0 and 1. To examine changes in learning rate across time, we assumed that `*A* is a person-specific parameter, given by the following growth function:

|  |  |  |
| --- | --- | --- |
|  |  | Equation 3 |

where `*Aij* is the (unbounded) learning rate for person *i* at occasion *j*, *β*0*i* is the person-specific intercept, *β*1*i* is the person-specific slope effect of time, *Tij* is the variable representing time, and *rij* is the residual. Residuals are assumed to be normally distributed with a mean of 0 and a variance of *σ*2, such that . Equation 3 represents level 1 of a hierarchical growth model in which the *β*s can be further decomposed into their group- and person-level effects at level 2, given by the following:

|  |  |  |
| --- | --- | --- |
|  |  | Equation 4 |

where *γ*00 and *γ*10 are the group-level intercept and slope, respectively, and *u0i* and *u1i* are the person-specific residuals from the group-level intercepts and slopes, respectively. Person-specific residuals are assumed to be multivariate normally distributed with means of 0, and a variance-covariance matrix given by the intercept and slope variances and , respectively, and covariance , such that

|  |  |  |
| --- | --- | --- |
|  |  |  |

|  |  |  |
| --- | --- | --- |
|  | No cor | Moderate cor |
| No effect | *rtime* = 0, *d* = 0 | *rtime* = .3, *d* = 0 |
| Moderate effect | *rtime* = 0, *d* = .5 | *rtime* = .3, *d* = .5 |

The four conditions represent parametric combinations of two levels of test-retest reliability, unreliable (i.e., *r* = 0) and moderate reliability (i.e., *r* = .3) and two levels of longitudinal change, no change (i.e., *d* = 0) and moderate change (i.e., *d* = .5; Cohen, 2016). Finally, after simulating data, we examined how well parameters could be recovered using more conventional (e.g., two-stage) approaches for analyzing longitudinal data using computational modeling.

## 3 Longitudinal Model of Iowa Gambling Task

The Iowa Gambling Task is a decision-making task that has been used in clinical populations to identify how individuals with various forms of psychopathology show unique patterns of decision-making compared to those without the that form of psychopathology. We chose this task to illustrate how adapt a computational model to a growth modeling framework because the IGT has a long history of being used to assess decision-making in a wide range of populations, including adolescent, adult, and clinical populations (e.g., individuals with anxiety, depression, & substance use disorders), and there have been multiple computational modeling built specifically for the IGT to assess decision-making among these populations.

## 4 Discussion

### 4.1 Benefits of this approach

### 4.2 Drawbacks of this approach

# References

# Tables

# Figures