Psychological Methods

Space-Time Modeling of Intensive Binary Time Series Eye-Tracking Data Using a Generalized Additive Logistic Regression Model

Sun-Joo Cho, Sarah Brown-Schmidt, Paul De Boeck, and Matthew Naveiras

Online First Publication, April 21, 2022. http://dx.doi.org/10.1037/met0000444

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Methods

ISSN: 1082-989X

https://doi.org/10.1037/met0000444

Space-Time Modeling of Intensive Binary Time Series Eye-Tracking Data Using a Generalized Additive Logistic Regression Model

Sun-Joo Cho[1], Sarah Brown-Schmidt[1], Paul De Boeck[2, 3], and Matthew Naveiras[1]

[1] Department of Psychology and Human Development, Peabody College, Vanderbilt University

[2] Department of Psychology, The Ohio State University

[3] Department of Psychology, KU Leuven

Abstract

Eye-tracking has emerged as a popular method for empirical studies of cognitive processes across multiple substantive research areas. Eye-tracking systems are capable of automatically generating fixation-location data over time at high temporal resolution. Often, the researcher obtains a binary measure of whether or not, at each point in time, the participant is fixating on a critical interest area or object in the real world or in a computerized display. Eye-tracking data are characterized by spatial-temporal correlations and random variability, driven by multiple fine-grained observations taken over small time intervals (e.g., every 10 ms). Ignoring these data complexities leads to biased inferences for the covariates of interest such as experimental condition effects. This article presents a novel application of a generalized additive logistic regression model for intensive binary time series eye-tracking data from a between- and within-subjects experimental design. The model is formulated as a generalized additive mixed model (GAMM) and implemented in the `mgcv` `R` package. The generalized additive logistic regression model was illustrated using an empirical data set aimed at understanding the accommodation of regional accents in spoken language processing. Accuracy of parameter estimates and the importance of modeling the spatial-temporal correlations in detecting the experimental condition effects were shown in conditions similar to our empirical data set via a simulation study.

Translational Abstract

A common technique for studying cognitive processes is the use of eye-tracking technology to monitor where the eyes are looking as a participant completes a task. Eye-tracking systems can generate data that is both spatially and temporally precise, and that can be used to make inferences about the underlying cognitive processes that support the task at hand. A common way of examining eye-tracking data is to obtain a measure of where, at each moment in time, the participant was looking. When the researcher is interested in a single target interest area, the researcher is working with intensive binary time-series data indicating whether or not, at each moment in time, the participant was fixating the target. In such data structures, we can expect strong temporal autocorrelation, as well as spatial-temporal correlations, because at certain time points, the distance between the current fixation position and the target location, and potential nontarget lures, may predict whether an upcoming target fixation is likely. The present article introduces a novel application of a generalized additive logistic regression model to intensive binary time series eye-tracking data, with typical data complexities as encountered in experimental studies of cognitive processes (i.e., experimental designs with within- and between-subjects factors, and crossed random effects). The model is applied to a dataset concerning spoken language processing, and is implemented in the `mgcv` `R` package. We show that modeling these spatial-temporal correlations is important for accurate statistical inference regarding fixed experimental condition effects.

Keywords: eye tracking, generalized additive mixed model, intensive binary time series data, logistic model, spatial-temporal correlations

Empirical studies of cognitive processes are used across multiple substantive research areas to explain the mechanisms that underlie fundamental processes such as perception, memory, language, learning, and decision making in psychology. These cognitive processes unfold over time and, as a result, researchers often leverage time-sensitive measures such as event-related brain

Funding was provided by the National Science Foundation (SES 1851690) to Sun-Joo Cho, Sarah Brown-Schmidt, and Paul De Boeck. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Illustrative data collection was supported by National Institute on Deafness and Other Communication Disorders (R01 DC011755) to Melissa Duff and Sarah Brown-Schmidt.

Correspondence concerning this article should be addressed to Sun-Joo Cho, Department of Psychology and Human Development, Peabody College, Vanderbilt University, 230 Appleton Place, Nashville, TN 37203, United States. Email: sj.cho@vanderbilt.edu

potentials (e.g., Kutas & Federmeier, 2011), reading times (e.g., MacDonald et al., 1994), and eye gaze during reading and other tasks (e.g., Tanenhaus et al., 1995; Warren et al., 2009). In particular, the study of eye gaze as participants complete a task with real world objects or a computerized display has emerged as a dominant method for examining various cognitive phenomena including visual memory (e.g., Hayhoe et al., 2003), problem solving (e.g., Thomas & Lleras, 2009), reading (e.g., Warren et al., 2009), children's word learning (e.g., Creel, 2014), and developmental research (e.g., Oakes, 2012). The resulting data exhibit a rich structure with multiple types of complexities that current data analytic approaches do not fully account for. The aim of the present work is to improve analytic techniques for eye-tracking data.

Using analytic techniques to process data is important for two major reasons. First, analytic techniques allow for a finer level of detail when capturing the ongoing processes of interest. These processes are complicated, and can change quickly over time in nonlinear ways. For analytic techniques to pick up on the features (i.e., characteristics) of these processes and to capture their complex nature with precision, the process data needs to be time-intensive, and the analytic methods need to be sufficiently flexible and detailed. Second, experimental studies are often used to test theories involving a hypothesized effect on cognitive processes and their outcomes. To test these effects in an accurate way, it is important to have accurate quantifications of the uncertainty of the estimated effects. The standard error of estimation does not simply depend on sample size, but it can be biased due to independence assumptions not being fulfilled. Time intensive data often show temporal (or serial) correlations[1] and trends, and the dynamic nature of the processes of interest implies that these correlations and trends may change over time. For eye-tracking data, there is also a spatial aspect of the data, which may give rise to spatial correlations. If these spatial and temporal aspects are not included in the analysis model, one would not be able to capture important aspects of the ongoing processes, and furthermore, the standard errors of the effects one wants to test may not be accurate. An additional challenge we are confronted with is that eye-tracking data are often coded as binary. Eye fixations can be registered in an (almost) continuous space, but when one is interested in gaze at particular objects, the eye-tracking data are categorical, indicating which of several candidates the participant is gazing at. If the researcher is interested in fixations to a specific (target) object, the data are thus coded as binary (i.e., 1 for target fixation; 0 for nontarget fixation). A binary coding scheme of this sort would be typical for cognitive tasks with a correct response option.

Spatial-Temporal Correlations in Intensive Binary Eye-Tracking Data

The temporal resolution of modern eye-trackers ranges from ~30 to 2,000 Hz depending on the eye-tracker model. The spatial resolution is expressed in terms of degrees of visual angle. For example, the data that we model in this article were recorded at 1,000 Hz and with a spatial resolution up to .1° visual angle. The eye-tracking system generates information about where in space the eye-tracked participant is looking at each moment in time. This information can then be used to ask questions such as how quickly and easily a participant understands an instruction or completes a task. For example, imagine we are studying how easily people follow instructions. To address this research question, participants are asked to view a computer screen with four different buttons and then told to "Push the green button to continue." As the participant listens to this instruction, we can generate a binary measure of whether or not the participant is fixating on the green button at each time point. In a task like this, we would expect participants to look for the green button (the "target") and then push it. As a result, whereas participants would initially be looking at the target button at chance levels, eventually most participants will have located the target button and thus fixations to the target will be common. This change will result in a nonlinear increase in the probability of fixating on the target over time. In addition to trend effects, temporal autocorrelations (AR; correlation of a variable with itself through time) from closely and equally time-spaced eye-tracking data are expected. Once the eyes move to a new fixation location, they are likely to linger at that new location as the participant comprehends the newly fixated information. As a result, with fast sampling rates, it is expected that the fixation position at prior time points is a strong predictor of fixation position at the current time point. This AR is expected to be highest at the beginning of the analysis time-window (before participants have had a chance to move their gaze) and at the end of the analysis time-window (once the participants have fixated on the target). The temporal AR may be smaller in the middle of the time window, when participants are likely to shift fixation from a nontarget button to the target button. Furthermore, all participants are presented with a sequence of *trials*. A trial is a presentation of an item. Because the same item can be presented multiple times, there can be more trials than items. As will be explained, the number of trials can differ across participants.

In addition to temporal trend and AR, spatial correlation is expected as well. Consider that the green target button extends over an area of space defined by the x, y coordinate boundaries of the button. When the participant's gaze is within the boundaries of the target area, the binary fixation data are coded as 1 (target fixation), and when they are outside of the target area the data are coded as 0 (nontarget fixation). In this example, nontarget fixations may include fixations to one of the other buttons on the screen, fixations to blank areas, or offscreen fixations. This type of data is expected to exhibit spatial correlations because all fixations within the area delimited by the x, y coordinates of the target will be coded as 1, and all other fixations will be coded as 0. In addition, due to the fact that the x, y coordinate location of an eye fixation changes over time when a participant moves their gaze, we also expect time-varying spatial effects. When the gaze fixates on a particular location, there is a functional field of view around that fixation location (Irwin, 2004). Following an initial saccade, participants may explore the area or correct the fixation landing position with secondary, corrective, or micro saccades (McCamy et al., 2014; Viviani & Swensson, 1982; Wu et al., 2010). As a result, we expect that the distance between the previous fixation point and a given interest area will predict the probability of fixating on that interest area in the current fixation point.

[1] We use the term of *correlation* to refer to any statistical association, although it commonly refers to the degree to which a pair of variables are linearly related.

Limitations and Strengths of Current Data Analysis Methods

Different kinds of statistical methods have been applied to eye-tracking data to answer research questions regarding visual attention while completing cognitive tasks as described above. Some researchers have used a *t*-test or a linear regression model based on an aggregate of binary eye-tracking data across a time window (Hanna & Brennan, 2007; Trude & Brown-Schmidt, 2012). Others have employed cluster randomization approaches in which a *t*-test or a multilevel logistic regression model is conducted at each time point to identify the time point at which proportions to an area of interest (AOI) differ between groups or experimental conditions (e.g., Barr et al., 2014; Oakes et al., 2013). These existing statistical methods are typically based on aggregated measures (i.e., proportions or logistic-transformed proportions) and therefore do not allow researchers to model spatial and temporal correlations and all sources of variability in the data. Failure to correctly model spatial and temporal correlations and all sources of variability can result in biased estimates for the effects of interest and biased standard errors. For example, Cho et al. (2018) showed that ignoring temporal AR in binary time-series eye-tracking data led to biased estimates for experimental condition effects and their standard errors.

Models that incorporate temporal AR have become increasingly popular in psychological research. For instance, the temporal AR in *continuous* time data has been modeled using multilevel linear models (e.g., de Haan-Rietdijk et al., 2016; Schuurman et al., 2016) and by using a state-space model (e.g., Oravecz et al., 2011) in affect research.

For binary (fixation vs. nonfixation) eye-tracking data, the temporal AR is modeled using a generalized linear mixed-effects model (GLMM; Cho et al., 2018; Cho, Brown-Schmidt, De Boeck, et al., 2020). However, previous studies did not consider spatial correlations in eye-tracking data. Although considerable work has been done on the visualization of spatial-temporal data from eye-trackers (e.g., Yarbus, 1967; Wooding, 2002), statistical modeling for such data is still very preliminary (see Cho, Brown-Schmidt, Naveiras, et al., 2020; Nixon et al., 2016; for exceptions). This article is among the first to model the spatial-temporal correlations in eye-tracking data in psychology.

Generalized additive models (GAM; Hastie & Tibshirani, 1990; Wood, 2017) have been used to model nonlinear temporal AR (e.g., Bringmann et al., 2017, 2018); nonlinear trend (Sullivan et al., 2015), nonlinear temporal AR and trend (Craigmile et al., 2010), and spatial-temporal correlations (Fang & Chan, 2015). The GAM was extended to the generalized additive mixed model (GAMM; Fahrmeir & Lang, 2001; Lin & Zhang, 1999; Wood, 2017) which permits the response probability distribution to be any member of the exponential family of distributions (e.g., Bernoulli, Poisson, Gamma). GAMM provides a flexible modeling framework to model data complexities from eye-tracking data. As with the GLMM, the GAMM allows for modeling variability, correlations between observations, and temporal or spatial correlations. A unique feature of the GAMM is that nonlinear dependencies between an outcome variable and continuous covariate(s) can be modeled using smooth functions. For example, temporal trend effects can be modeled in a data-driven manner using GAMM, instead of a parametric (e.g., linear or quadratic) trend. However, GAMM has not been used previously to answer research questions based on intensive binary time series data with space and time related complexities.

Study Purpose

The purpose of this article is to illustrate how one can account for temporal and spatial dynamics in eye-tracking data using a *generalized additive logistic regression model* as a GAMM, so that the effects of interest can be tested in a more accurate way. In the generalized additive logistic regression model used here, smooth functions are specified to model the following types of possible nonlinearities:

- The level of the outcome variable (i.e., target fixations) may change in a nonlinear way across time points within trials and across the sequence of trials. The former is a trend within the time series and the latter is a trend across the time series.
- The trend within the time series may depend on the location of the time series in the sequence of trials. The interaction between trial and trend within a time series may be nonlinear.
- Temporal AR and correlations with the location of the previous observation (i.e., the previous fixation) may change in a nonlinear way within time series.
- The spatial correlations in the outcome variable from the *x*, *y* coordinates of a computer screen are expected. These spatial correlations are about the concentrations of eye fixations.

Specifying these spatial and temporal aspects of the data in a generalized additive logistic regression model leads to more accurate statistical inference for the hypothesized effects of interest referring to the factors in an experiment. To the best of our knowledge, the combination of these four effects listed have not been considered in the application of GAMM for intensive binary time series data. In addition, parametric random effects in GAMM were also considered to take the variability of persons and items into account, as is recommended for psycholinguistic data and as is commonly implemented in GLMM (Baayen et al., 2008). We utilize the `mgcv` package Version 1.8–28 (Wood, 2019) in `R` Version 3.2.4 (R Core Team, 2016) for estimation of the generalized additive logistic regression model.

This article is organized as follows. First, we explain how trend and spatial-temporal correlations can be explored graphically and descriptively. Second, we describe the smooth additive logistic regression model and explain the estimation, model evaluation, and testing in the `mgcv` package. Third, we illustrate the model applied to the empirical data set from Trude et al. (2013). Subsequently, we provide a simulation study to investigate the accuracy of parameter estimates and standard errors, and the consequences of ignoring the spatial-temporal correlations in detecting experimental condition effects. Finally, we end with a summary and a discussion.

Characterizing Time, Space, and Related Correlations

A typical data structure for a visual world eye-tracking study is a multilevel structure with three levels for binary data (Barr,

2008). Specifically, time point data at Level 1 are nested within trials at Level 2 (one time series per trial), which are cross-classified by persons and items at Level 3. The data typically come from more than one trial, which means that each person has time series data for multiple items across multiple trials, and may experience the same item multiple times on different trials. In this multilevel data structure, an individual trial (which takes place over a sequence of time points) constitutes a time series. As explained in the Introduction, it is expected that there are change processes (trend and temporal AR), spatial correlations, and spatial AR in the time series for eye-tracking data. It is a common practice to begin with an exploratory descriptive analysis of time series and spatial data (e.g., Cressie & Wikle, 2011). Below, descriptive methods are provided as a preliminary analysis.

Trend and Temporal AR

A time-series plot (with time on the *x*-axis and the outcome variable on the *y*-axis) can be created to explore the trend over time. In addition, autocorrelations (Box & Jenkins, 1976) can be calculated to investigate whether there is a trend or AR or both. In the presence of both trend and AR, the autocorrelations for small lags tend to be large and positive because observations nearby in time are also close by in value (Chatfield, 2004). Furthermore, partial autocorrelations can be calculated to select the order of autocorrelations (Chatfield, 2004). In this study, autocorrelations and partial autocorrelations were calculated using the `acf` and `pacf` functions respectively in the `stats` `R` package (R Core Team, 2016).

Spatial Correlations

In eye-tracking data measured using the *x*, *y* coordinates of eye fixations on a computer screen, it is expected that eye fixations close to each other are likely to be more similar than eye fixations that are further apart. To investigate whether such spatial correlations exist in eye-tracking data, a *variogram* (Cressie, 1993) can be calculated. Similar to how temporal autocorrelations measure temporal dependence by comparing values at time points t and $(t + lag)$, the variogram measures spatial dependence by comparing values in space (i.e., the *x*, *y* coordinates on a computer screen) with the distances that separate those points. A low value of the variogram indicates spatial dependence (e.g., eye fixations in close proximity), whereas a large value indicates spatial independence (e.g., eye fixations far apart).

Let y_{tlji} for trial l ($l = 1, \ldots, L$), person j ($j = 1, \ldots, J$), and item i ($i = 1, \ldots, I$) at time point t ($t = 1, \ldots T$ for equally-spaced time points) denote a binary-coded fixation data point, coded 1 ($y_{tlji} = 1$) for a predefined fixation (i.e., looking at an image), and coded 0 ($y_{tlji} = 0$) otherwise. The variogram of residuals from a *null* model can be investigated prior to modeling the spatial correlation (e.g., Zuur et al., 2009). The motivating data in this article come from the field of psycholinguistics, where simultaneously modeling person and item variability using crossed random person and item effects has been widely advocated (Baayen et al., 2008; Barr, 2008; Jaeger, 2008; Quené & van den Bergh, 2008). Thus, the null model can include a fixed intercept parameter δ_0, a (parametric) random person effect θ_j, and a (parametric) random item effect ζ_i, as follows:

$$log\frac{P(y_{tlji}=1\mid\theta_j,\zeta_i)}{1-P(y_{tlji}=1\mid\theta_j,\zeta_i)}=\text{logit}[P(y_{tlji}=1\mid\theta_j,\zeta_i)]=\delta_0+\theta_j+\zeta_i. \quad (1)$$

The null model is a Rasch, or 1-parameter logistic, model in which a person parameter (a latent variable) and an item parameter (item location) are random effects (De Boeck, 2008). As will be shown later, the null model can also be formulated as a GLMM with crossed person and item random effects (e.g., Baayen et al., 2008). The person random effect is for modeling individual differences in target fixations, and the item random effect is for modeling variability in item location across items. The Pearson residuals of the null model (denoted by R) are used to calculate the variogram.

To explore spatial dependence, the variogram can be plotted for points separated by the same euclidean distance Q. Denote the residuals R of two data points with Q distance between them by $R(n, m)$ and $R(n + d_n, m + d_m)$, with distance d_n between the x coordinates and distance d_m between the y coordinates such that $\sqrt{d_n^2 + d_m^2} = Q$. Figure A1 of Appendix A illustrates the distance Q between the x, y coordinates of $R(n, m)$ and $R(n + d_n, m + d_m)$. In practice, it is difficult to find enough data points separated by exactly the same distance Q. Thus, the set of all possible distances Q is partitioned into *classes*. As an example, Figure A2 of Appendix A presents four classes separated by "Distance 1," "Distance 2," "Distance 3," and "Distance 4." Data points within each ring in Figure A2 of Appendix A are grouped into one class. For each class, the variogram is calculated as the variance of the differences of residuals among pairs of data points:

$$\hat{\gamma}(Q_c)=\frac{1}{N_c}\sum_{q=1}^{N_c}\{R(n_q+d_{n,q},m_q+d_{m,q})-R(n_q,m_q)\}^2, \quad (2)$$

where q is an index for a pair of data points within class Q_c ($c = 1, \ldots, C$), and N_c is the number of pairs of data points within class Q_c. In practice, a semivariogram (i.e., the semivariance) is used, which is equal to half the value of the variogram ($0.5\hat{\gamma}(Q_c)$). A semivariogram includes the variance for each pair of points once, rather than once for each point within that pair (e.g., Webster & Oliver, 2007). The calculated semivariogram can be plotted for all classes against the distance to observe patterns in spatial dependence. The expected pattern in the plot is that the value of the semivariogram is zero for a distance of zero, and that increasing the distance increases the value of the semivariogram. Such a plot would indicate that a decrease in spatial dependence (i.e., an increase in the distance between points) is associated with an increase in the value of the semivariogram (i.e., an increase in the variance of the difference of residuals; see Figure A3 of Appendix A). In Appendix A, the calculation of the semivariogram is illustrated in R using 40 observations.

Data points from all possible locations within a class (i.e., within the ring associated with that class, see Figure A2 in Appendix A) are used for calculating the semivariogram using Equation 2, with the assumption that the spatial dependence of the residuals is the same in every direction (e.g., 0° [north], 45° [northeast], 90° [east], and 135° [southeast], see Figure A4 of Appendix A for different directions of x, y coordinates). This assumption, called

isotropy, can be tested by calculating the semivariogram for each direction (e.g., four semivariograms, one for each direction of north, northeast, east, and southeast). Figure A4 of Appendix A illustrates this test of isotropy, separating the data to calculate the variogram using the northeast (red dots) data points and then again using the southeast (blue dots) data points. The isotropy assumption is reasonable when the strength and pattern of the semivariogram is similar for each direction. The `variogram` function in the `gstat` `R` package (Pebesma, 2004) is used in this study.

Spatial by Temporal Interactions

There are a few different ways to characterize spatial by temporal interactions (see Cressie & Wikle, 2011 for a review). First, spatial correlations can be presented using a semivariogram at different time points. When there are many time points, and each with a small time interval (e.g., 132 time points at 10-ms intervals), time points can be grouped into blocks (e.g., six time blocks; each consisting of 22 time points of 10 ms) for calculating the semivariogram. Second, patterns in x or y coordinates over time can be investigated, that is, x coordinate versus time and y coordinate versus time. Third, temporal correlations can be investigated within a location in a grid of the x, y coordinates on a computer screen. In this study, we chose the first approach (i.e., spatial correlations by time blocks) to show the spatial-by-temporal interaction because we are interested in exploring whether the effect of spatial information (e.g., distance from a fixation to a target location or a competitor location) differs over time.

The Generalized Additive Logistic Regression Model

In this section, existing GAMM applications are first reviewed to highlight the limitations of the current methods of analyzing binary time series eye-tracking data. Then, the generalized additive logistic regression model is specified as a GAMM. Subsequently, the estimation, model evaluation, and testing are explained in the `mgcv` package.

Existing GAMM Applications to Time Series Data

GAMMs have been applied to different kinds of time series: pupil dilation data (e.g., Loo et al., 2016; Vogelzang et al., 2016), event-related brain potential (ERP) data (e.g., Tremblay & Newman, 2015), reaction time (RT; e.g., Baayen, 2010; Baayen et al., 2017); and educational intervention data from single case designs (Shadish et al., 2014). Furthermore, there are recent GAMM applications in which temporal AR was also modeled for continuous time series data (with an identity link; e.g., Baayen et al., 2018; Baayen et al., 2017; Wieling et al., 2016). GAMMs have also been applied to continuous time-series pupilometric data with temporal AR in the residuals and spatial gaze information, with x, y coordinates of the gaze position on screen as a covariate (van Rij et al., 2019). Related to GAMM, a linear spline model as a mixed model was applied to model a nonlinear trend over time and temporal AR of the residuals, based on an aggregated measure (proportion of time that participants were looking at the target) from eye-tracking data (Yamashiro et al., 2019). In addition to using gaze position to model spatial correlation as a covariate as in van Rij et al. (2019), the *distance* between the fixation point and the centroid of the target interest area can be considered as another kind of spatial information. Nixon et al. (2016) considered GAMM for eye-tracking data and the Euclidean distance between the current fixation position and the centroid of the target and competitor pictures as continuous outcome variables (instead of as covariates) in the GAMM. Cho, Brown-Schmidt, Naveiras, et al. (2020) considered fixed spatial lag effects in a model which showed the probability of a target fixation at a time point t is higher when the fixation at the previous time point $t-1$ is close to the target. More flexible spatial lag effects can be modeled such as nonlinear spatial lag and time-varying spatial lag effects (as covariates) in the GAMM, which has not been illustrated before.

Except for Cho, Brown-Schmidt, Naveiras, et al. (2020), all the aforementioned applications have in common that the outcome variable data are continuous so that the model specifications are not applicable to binary data. For binary data, a correlated residuals approach is difficult to implement (Cox & Snell, 1989). Instead, it is more common to model AR using the regression of current outcomes on past outcomes (e.g., Bartolucci & Nigro, 2010; Cox & Snell, 1989; Fokianos & Kedem, 2003, Chapter 2; Hung et al., 2008; Zeger & Qaqish, 1988). Cox and Snell (1989, pp. 100–101) noted that this approach is *observation driven* in the sense that the probability at *time t* is determined by the observations at previous time points. The observation driven approach is appealing because it is easy to fit a GAMM without further development of estimation algorithms and inference, so that one can rely on the `mgcv` package.

Generalized Additive Logistic Regression Model

Below, we present GLMM as a model related to GAMM and then introduce a general form of GAMM (e.g., Wood, 2017). Subsequently, the generalized additive logistic regression model is specified as GAMM for space-time modeling of intensive binary time series eye-tracking data.

GLMM and GAMM

GLMM (e.g., Laird & Ware, 1982) is specified for an outcome variable y distributed as an exponential family distribution (*EF*; e.g., Bernoulli, Poisson, Gamma distributions) with mean μ and scale parameter ϕ:

$$y \sim EF(\mu, \phi). \tag{3}$$

The mean μ is related to the linear predictor $\mathbf{X}'\boldsymbol{\delta} + \mathbf{Z}'\mathbf{b}$ with a link function g:

$$g(\mu) = \mathbf{X}'\boldsymbol{\delta} + \mathbf{Z}'\mathbf{b}, \tag{4}$$

where $\mathbf{X}$ is a design matrix for fixed effects, $\mathbf{Z}$ is a design matrix for random effects, $\boldsymbol{\delta}$ is the vector of fixed parameters, and $\mathbf{b}$ is the vector of random parameters. The vector of random parameters $\mathbf{b}$ is assumed to follow a multivariate normal (*MVN*) distribution with a mean vector of $\mathbf{0}$ and a variance-covariance matrix $\boldsymbol{\Sigma}$, $\mathbf{b} \sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$.

The null model presented in Equation 1 can be presented as GLMM as follows:

$$\text{logit}[P(y_{tlji} = 1 \mid \theta_j, \zeta_i)] = x'\delta_0 + z_j'\theta_j + z_i'\zeta_i, \tag{5}$$

where x' is a design matrix for a fixed intercept (a vector of 1s), z_j is a design matrix for a parametric random person effect (a vector of 1s for a random person intercept), and z_i is a design matrix for a

parametric random item effect (a vector of 1s for a random item intercept).

GAMM (e.g., Wood, 2017) is a GLMM in which the linear predictor partly depends linearly on some unknown smooth functions (f_h) of the covariate(s):

$$g(\mu) = \mathbf{X}'\boldsymbol{\delta} + \mathbf{Z}'\mathbf{b} + \sum_{h=1}^{h'} f_h(x_h) + \sum_{h=h'+1}^{H} f_h(x_{1h}, x_{2h}), \tag{6}$$

where h is an index for the smooth function ($h = 1, \ldots, h'$ for univariate smooth functions; $h = h' + 1, \ldots, H$ for bivariate smooth functions), $f_h(x_h)$ is a univariate smooth function of a covariate x_h, and $f_h(x_{1h}, x_{2h})$ is a bivariate smooth function of the two covariates x_{1h}, x_{2h}.

The univariate smooth function $f_h(x_h)$ of a covariate x_h is the weighted sum of a set of basis functions defined over the covariate x_h:

$$f_h(x_h) = \sum_{k=1}^{K} \gamma_{hk} b_{hk}(x_h), \tag{7}$$

where k is an index for a basis function ($k = 1, \ldots, K$), x_h is a covariate for a smooth function h, γ_{hk} is the kth basis coefficient, and $b_{hk}(x)$ is the kth basis function for smooth function h.

The bivariate smooth function of the two covariates (x_{1h} and x_{2h}) is the weighted sum of a set of basis functions defined over the covariates:

$$f_h(x_{1h}, x_{2h}) = \sum_{k}^{K} \sum_{k'}^{K'} \gamma_{hkk'} b_{hkk'}(x_{1h}, x_{2h}), \tag{8}$$

where k is an index for a basis function ($k = 1, \ldots, K$) for a covariate x_{1h}, k' is an index for a basis function ($k' = 1, \ldots, K'$) for a covariate x_{2h}, $\gamma_{hkk'}$ is a basis coefficient, and $b_{hkk'}(x_{1h}, x_{2h})$ is a bivariate basis function. For a tensor product smooth as an example, the bivariate basis function is $b_{hkk'}(x_{1h}, x_{2h}) = b_{hk}(x_{1h}) b_{hk'}(x_{2h})$, which is all pairwise products of univariate basis functions in x_{1h} and x_{2h} directions.

Generalized Additive Logistic Regression Model as GAMM

Binary time series eye-tracking data y_{tlji} (for time point t, trial l, person j, and item i) have an independent Bernoulli distribution with mean μ and variance $\mu(1 - \mu)$:

$$y_{tlji} \sim Bernoulli(\mu), \tag{9}$$

where μ is a probability of a fixation conditional on random person effect θ_j and random item effect ζ_i, $P(y_{tlji} = 1 \mid \theta_j, \zeta_i)$. The scale parameter ϕ in Equation 3 is equal to 1 for Bernoulli distributions (e.g., Wood, 2017, p. 104). The mean μ in the Bernoulli distribution is related to the linear predictor of GAMM with a logit link function:

$$\begin{aligned} log\frac{P(y_{tlji} = 1 \mid \theta_j, \zeta_i)}{1 - P(y_{tlji} = 1 \mid \theta_j, \zeta_i)} &= \text{logit}[P(y_{tlji} = 1 \mid \theta_j, \zeta_i)] \\ &= \mathbf{X}'\boldsymbol{\delta} + z_j'\theta_j + z_i'\zeta_i + \sum_{h=1}^{h'} f_h(x_h) + \sum_{h=h'+1}^{H} f_h(x_{1h}, x_{2h}). \end{aligned} \tag{10}$$

In the following, smooth functions are specified to model nonlinear spatial-temporal effects in eye-tracking data based on Equation 10. Table 1 illustrates the covariates of the smooth functions for the first 40 time points of data for a single combination of trial, person, and item, as an example.

Smooth Functions for Time and Trials

The following model includes smooth functions for time and trials to model processes in the eye-tracking data:

$$\begin{aligned} \text{logit}[P(y_{tlji} = 1 \mid \theta_j, \zeta_i)] &= \mathbf{X}'\boldsymbol{\delta} + z_j'\theta_j + z_i'\zeta_i \\ &+ f_1(time_t) + f_2(trial_l) + f_3(time_t, trial_l) + y'_{(t-1)lji} f_4(time_t), \end{aligned} \tag{11}$$

where

- $time_t$ is an equally-spaced time covariate,
- $trial_l$ is an equally-spaced trial covariate indicating the ordinal position of the trial (the same for all persons),
- $y_{(t-1)lji}$ is the first-order temporal lag covariate, which is created for a unique combination of a trial, a person, and an item. As shown in Table 1, $y_{(t-1)lji}$ allows the researcher to use the prior time point ($t - 1$) as a covariate of fixations at the current time point (t) in order to address the AR between time points that is commonly observed in eye-tracking data,
- $f_1(time_t)$ and $f_2(trial_l)$ are univariate smooth functions to model nonlinear and possibly even non-monotonic trends in the probability of target fixations over time within the time series and across trials, respectively,
- $f_3(time_t, trial_l)$ is a bivariate smooth function to model the interaction between time and trials, as in the analysis of variance (ANOVA) decomposition. The interaction of time and trials indicates that trend over time differs across trials, and
- $y'_{(t-1)lji} f_4(time_t)$ is a univariate smooth function of time for the first-order lag temporal lag covariate. The temporal AR effect is considered to model temporal autocorrelations between adjacent binary responses (Cox & Snell, 1989), which may vary over time. As explained earlier, it is expected that the temporal AR is the highest at the beginning and end of the time-window and is smaller in the middle of the time-window.

Smooth Functions for Time and Trials and Their Interactions With Space

We hypothesize that the probability of a target fixation is higher the closer the fixation at the previous time point $t - 1$ was to the target, and the further the fixation at $t - 1$ was from the competitor. Thus, spatial lag covariates were created to account for variation in the probability of target fixation. Adding smooth functions of these spatial lag covariates to Equation 11 leads to:

$$\begin{aligned} \text{logit}[P(y_{tlji} = 1 \mid \theta_j, \zeta_i)] &= \mathbf{X}'\boldsymbol{\delta} + z_j'\theta_j + z_i'\zeta_i \\ &+ f_1(time_t) + f_2(trial_l) + f_3(time_t, trial_l) + y'_{(t-1)lji} f_4(time_t) \\ &+ d'_{(t-1)lji} f_5(time_t) + c'_{(t-1)lji} f_6(time_t) \end{aligned} \tag{12}$$

where

- $d_{(t-1)lji}$ is the first-order spatial lag covariate for the target. In order to create the covariate, the Euclidean distance between the fixation point in terms of x, y coordinates of the

Table 1
An Illustration of Covariates: Data Over First 40 Time Points for a Trial, a Person, and an Item

			Locations						Euclidean distance		Lag covariates		
Time	Trial	y_{tlji}	$n_{f.tlji}$	$m_{f.tlji}$	n_v	m_v	$n_{v'}$	$m_{v'}$	D_{tlji}	C_{tlji}	$y_{(t-1)lji}$	$d_{(t-1)lji}$	$c_{(t-1)lji}$
1	1	0	1,247	137	690	270	1,230	270	572.66	134.08	—	—	—
2	1	0	1,247	137	690	270	1,230	270	572.66	134.08	0	572.66	134.08
3	1	0	1,247	137	690	270	1,230	270	572.66	134.08	0	572.66	134.08
4	1	0	1,247	137	690	270	1,230	270	572.66	134.08	0	572.66	134.08
5	1	0	1,247	137	690	270	1,230	270	572.66	134.08	0	572.66	134.08
6	1	0	1,247	137	690	270	1,230	270	572.66	134.08	0	572.66	134.08
7	1	0	1,247	137	690	270	1,230	270	572.66	134.08	0	572.66	134.08
8	1	0	1,247	137	690	270	1,230	270	572.66	134.08	0	572.66	134.08
9	1	1	678	150	690	270	1,230	270	120.60	564.89	0	572.66	134.08
10	1	1	678	150	690	270	1,230	270	120.60	564.89	1	120.60	564.89
11	1	1	678	150	690	270	1,230	270	120.60	564.89	1	120.60	564.89
12	1	1	678	150	690	270	1,230	270	120.60	564.89	1	120.60	564.89
13	1	1	678	150	690	270	1,230	270	120.60	564.89	1	120.60	564.89
14	1	1	678	150	690	270	1,230	270	120.60	564.89	1	120.60	564.89
15	1	1	690	146	690	270	1,230	270	124.00	554.05	1	120.60	564.89
16	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
17	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
18	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
19	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
20	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
21	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
22	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
23	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
24	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
25	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
26	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
27	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
28	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
29	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
30	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
31	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
32	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
33	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
34	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
35	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
36	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
37	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
38	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
39	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05
40	1	1	690	146	690	270	1,230	270	124.00	554.05	1	124.00	554.05

Note. – indicates a missing value; the $n_{f.tlji}$ and $m_{f.tlji}$ are the x and y coordinates for the fixation location f; the n_v and m_v are the x and y coordinates for the target location v; the $n_{v'}$ and $m_{v'}$ are the x and y coordinates for the competitor location v'.

computer screen $(n_{f.tlji}, m_{f.tlji})$ at time t and the centroid of the target location (n_v, m_v)[2], is first calculated as follows:

$$d_{tlji} = \sqrt{(n_v - n_{f.tlji})^2 + (m_v - m_{f.tlji})^2}. \quad (13)$$

Based on d_{tlji}, $d_{(t-1)lji}$ is obtained. To illustrate the calculation of d_{tlji} and $d_{(t-1)lji}$, $(n_{f.11ji}, m_{f.11ji}) = (1247, 137)$ and $(n_v, m_v) = (690, 270)$ from the first trial $l = 1$ at the first time point $t = 1$ are chosen from Table 1 and the Euclidean distance between the fixation point and the centroid of the target location are calculated as follows:

$$d_{11ji} = \sqrt{(690 - 1247)^2 + (270 - 137)^2} = 572.66. \quad (14)$$

The $d_{11ji} = 572.66$ is then used to obtain the first-order spatial lag covariate for the target at time point $t = 2$, $d_{11ji} = 572.66$.

- $c_{(t-1)lji}$ is the first-order spatial lag covariate for the competitor. The Euclidean distance between the fixation point in terms of x, y coordinates of the computer screen $(n_{f.tlji}, m_{f.tlji})$ at time t and the centroid of the competitor location $(n_{v'}, m_{v'})$[3], is first calculated as follows:

$$c_{tlji} = \sqrt{(n_{v'} - n_{f.tlji})^2 + (m_{v'} - m_{f.tlji})^2}. \quad (15)$$

The $c_{(t-1)lji}$ is then obtained using c_{tlji}. As an example, $(n_{f.11ji}, m_{f.11ji}) = (1247, 137)$ and $(n_v, m_v) = (1230, 270)$ from the first trial $l = 1$ at the first time point $t = 1$ are chosen from Table 1 and the Euclidean distance between the fixation point and the centroid of the competitor location are calculated as follows:

[2] Note that the n_v and m_v for a target is fixed per trial, person, and item.

[3] Note that the $n_{v'}$ and $m_{v'}$ for a competitor is fixed per trial, person, and item.

$$c_{11ji} = \sqrt{(1230 - 1247)^2 + (270 - 137)^2} = 134.08. \quad (16)$$

The first-order spatial lag covariate for the competitor at the second time point $t = 2$, c_{11ji}, is then 134.08, and

- $d'_{(t-1)lji} f_4(time_t)$ and $c'_{(t-1)lji} f_5(time_t)$ are univariate smooth functions of time to model time-varying effects of the first-order spatial lag covariates of target and competitor, respectively.

Smooth Functions for Time and Trials, Space, and the Interactions of Time and Trials With Space

To model spatial correlations in the binary fixation data, a bivariate smooth function for values on the x, y coordinates of the computer screen is added to Equation 12:

$$\begin{aligned}\text{logit}[P(y_{tlji} = 1 \mid \theta_j, \zeta_i)] &= \mathbf{X}'\boldsymbol{\delta} + z'_j\theta_j + z'_i\zeta_i \\ &+ f_1(time_t) + f_2(trial_l) + f_3(time_t, trial_l) + y'_{(t-1)lji} f_4(time_t) \\ &+ d'_{(t-1)lji} f_5(time_t) + c'_{(t-1)lji} f_6(time_t) \\ &+ f_7(xposition_{tlji}, yposition_{tlji}), \end{aligned} \quad (17)$$

where

- $(xposition_{tlji}, yposition_{tlji}) = (n_{f.tlji}, m_{f.tlji})$ are values of the x, y coordinates of the computer screen in which binary fixation data points are located and
- $f_7(xposition_{tlji}, yposition_{tlji})$ is a bivariate smooth function to model spatial correlations in the binary fixation data based on the x, y coordinates of a computer screen. The closer fixation points on a computer screen are over time, the higher the spatial correlations are.

In Equation 17, spatial-temporal effects (seven smooth functions) and random effects (θ_j and ζ_i) are controlling parameters for the fixed focal parameters (e.g., experimental condition effects).

Estimation

Below, we explain estimation details of smooth functions and model parameters in the `mgcv` package.

Smooth Functions in the `mgcv` Package

For the univariate smooth function ($f_h(x_h)$), a cubic regression spline (CRS; Wood, 2017) and a thin plate regression spline (TPRS; Wood, 2017, 5.5.1), which are commonly used splines in GAMM applications, are used with the `mgcv` `R` package. The CRS is a smooth curve comprised of sections of cubic polynomials in which the sections are joined together at some specified locations (called *knots*). At the knots, the two sections of the cubic polynomials that meet have the same value (otherwise, the function is not continuous), as well as the same first and second derivative (Wood, 2017, 5.3.1). The knot points are automatically placed (with equal spacing) over the entire range of the observed covariate by default in the `mgcv` package. To illustrate the CRS, an additive logistic model with one smooth function of a time covariate (x_t; $t = 1, \ldots, 132$) is considered using a binary time series with 132 time points with $K = 10$ (where K is the number of basis functions). Figure 1 (top left) illustrates the CRS basis function ($b_k(x_h)$) with 10 knots evenly distributed through the range of the time covariate (x_t). Figure 1 (left top) presents nine CRS basis expansions ($10 - 1$ basis functions due to model identification constraint, explained below shortly) spread evenly through the range of the time covariate (1.000, 15.556, 30.111, 44.667, 59.222, 73.778, 88.333, 102.889, 117.444, 132.000). In addition, Figure 1 (top right) shows the same CRS basis functions weighted by the estimated coefficients ($\hat{\gamma} = [4.127, 4.208, 3.429, 3.765, 3.082, 0.966, 1.991, 1.977, -0.278]'$) and the resulting trend line (dotted line in the figure) created by summing up the weighted basis function at the different values of x_t.

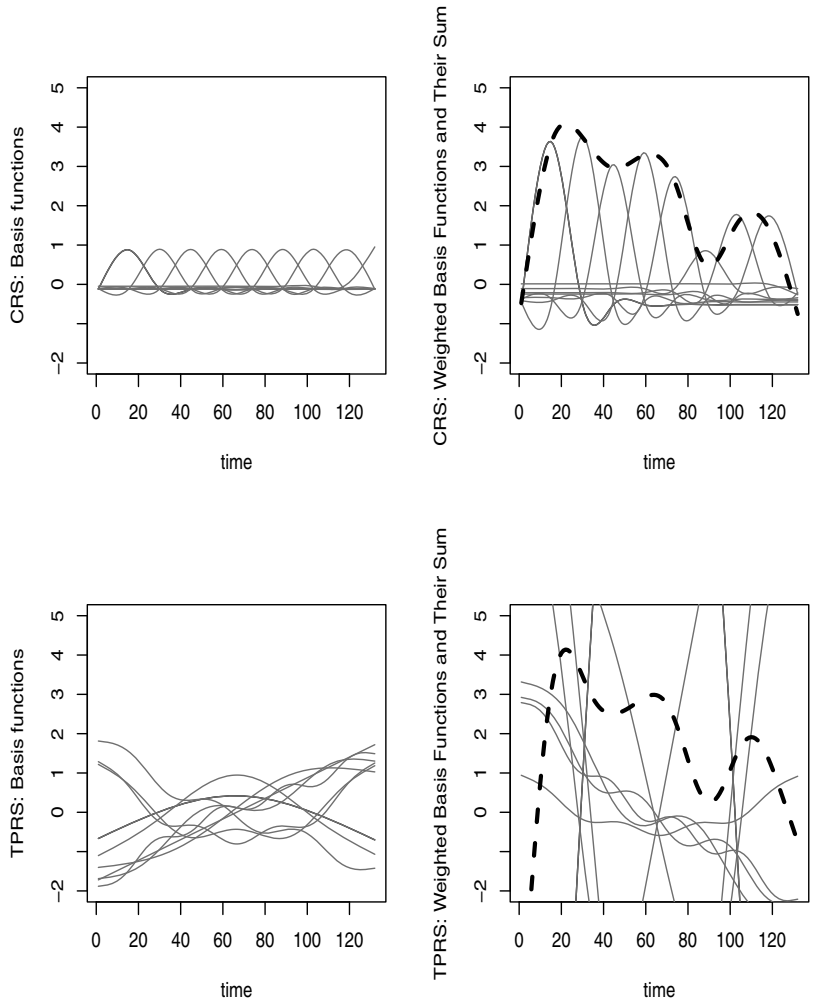
Unlike setting the knots in the CRS, the TPRS uses an eigen-decomposition of the basis functions to find the basis coefficients that maximally account for the variances in the data. Thus, arbitrarily choosing knot locations can be avoided. However, the eigen-decomposition and related steps can be costly for a large data set ($\sim 1,000$ observations). Figure 1 (bottom left) presents the TPRS basis function ($b_k(x)$) with $K = 10$ for the same data set used in the CRS illustration. Furthermore, Figure 1 (bottom right) shows the same TPRS basis functions weighted by the estimated coefficients ($\hat{\gamma} = [46.853, -5.779, -13.846, -1.964, 7.701, 1.613, 0.732, -1.483, 10.813]'$), with the resulting trend line (dotted line in the figure).

In the univariate smooth function, the TPRS provides very similar results to the CRS spline in applications (e.g., Finch & Finch, 2018), which is also illustrated in Figure 1. In this study, the CRS was chosen for the univariate and bivariate smooth functions (except for one bivariate smooth function as will be explained below) because of its computational efficiency for a large scale eye-tracking data set. However, we also considered the TPRS to check whether an alternative smoother provides the same results.

For bivariate smooth functions ($f_3(time_t, trial_l)$ and $f_7(xposition_{tlji}, yposition_{tlji})$ in Equation 17), different kinds of smoothing basis functions can be chosen depending on whether the variables are measured in the same units or in different units (Wood, 2017, p. 227). In a scale-invariant smoothing function such as a tensor product smooth, a unit change in one variable is equivalent to a unit change in another variable. Thus, the tensor product smooth can be used when the relative scaling of the covariates of the smooth is quantified in different units. Because $time_t$ and $trial_l$ are measured in different units, a tensor product smooth interaction was chosen for $f_3(time_t, trial_l)$. In contrast, the TPRS is an *isotropic* smooth function, which uses the same smoothness per unit change in the two covariates (x_{1h} and x_{2h}). Accordingly, the TPRS is appropriate when the relative scaling of the covariates of the smooth is naturally measured in the same units. Because the x, y coordinates of the gaze locations on a screen are expressed in the same units (in pixels), the TPRS was chosen for $f_7(xposition_{tlji}, yposition_{tlji})$. In Appendix B, we present how random effects and smooth functions in Equation 17 (a model having all smooth functions considered) can be specified in the `mgcv` package.

In the `mgcv` package, a model is estimated with an identification constraint that the function f_h sum over the observed covariate values is 0 (i.e., $\sum_v f_h(x_{hv}) = \mathbf{0}$ for each h with v as a subscript for observations; called a sum-to-zero constraint). The

Figure 1
Illustrations of Basis and Scaled Functions



Note. Basis functions of the CRS with knots spread evenly through the range of a time covariate (top left), the scaled CRS functions weighted by the estimated coefficients γ with the resulting trend line (dotted line; top right), basis functions of the TPRS with knots spread evenly through the range of a time covariate (bottom left), and the scaled TPRS functions weighted by the estimated coefficients ***\gamma*** with the resulting trend line (dotted line; bottom right).

identification constraint is set by estimating $K-1$ basis coefficients with $K-1$ basis functions for each smooth function (where K is the number of basis functions in Equation 7). The result of this constraint is that the (fixed) intercept in the model is the mean of smooth functions in the model. To illustrate the sum-to-zero constraint, a data set is generated under a simple logistic model with a smooth function of a covariate x_1 ($\text{logit}[P(y=1)] = f_1(x_1) = sin(x_1+2)$) for 5,000 observations in R:

```
library(mgcv)
set.seed(1242)
#generate a covariate of x1
x1 <- runif(5000,0,1)
#generate a smooth function of f1 using a sin
  #function; mean(f1)=intercept
f1 <- sin(x1 + 2)
mean.f1 <- mean(f1)
mean.f1
> mean.f1
[1] 0.5699139
#generate the probability of a response
  #equaling 1
probability <- 1/(1+exp(-f1))
#generate Bernoulli response variable
y <- rbinom(5000,1,probability)
#creating a data set
data <- data.frame(x1,y)
#fit a logistic model using the bam function
  #with K = 4
model <- bam(y ~ 1 + s(x1,k = 4),data=data,
  family=binomial,method="ML")
```

```
#extract intercept and basis coefficient
  #estimates
gamma <- model$coefficients
gamma
> gamma
```

(Intercept)	s(x1).1	s(x1).2	s(x1).3
5.311421e-01	−6.570583e-04	−6.570583e-04	−2.121355e-01

The mean of the generated smooth function is 0.5699139, which is estimated as an intercept estimate, $5.311421e-01 = 0.5311421$, as shown above. Three basis coefficients (for three basis functions; $K - 1 = 3$ where $K = 4$ in estimation due to the sum-to-zero constraint) are estimated as $[-6.570583e-04, -6.570583e-04, -2.121355e-01]'$. For the bivariate smooth functions of (x_{1h} and x_{2h}), the sum-to-zero constraint is imposed for each covariate prior to constructing the tensor product basis.

For the selected basis functions for univariate and bivariate smooths, the number of basis functions (K) should be selected to obtain a good fit. The role of K is to set the dimensionality of the basis expansion. Oversmoothing is expected when the number of basis functions is too small. Wiggly curves and slower computation time are expected when the number of basis functions is too large. In this study, we set $K = 10$ (default in `mgcv`) for all smooth functions. One way to check whether a selected K is large enough is to examine the value of the k-index (which is provided in the output from the gam.check function in `mgcv`). If the k-index (see Wood, 2017, p. 330 for the technical details) is lower than 1, a larger number of K should be considered.

In the `mgcv` package, a measure of nonlinearity is provided using the effective degrees of freedom (*edf*). The *edf* is an estimate of the degrees of freedom that are used by a smooth with a given number of basis functions and a given smoothing parameter. The higher the *edf*, the more wiggly the estimated smooth function is. An *edf* of 1 indicates a linear effect of a covariate on the outcome, an *edf* of 2 indicates an approximately quadratic effect of a covariate on the outcome, and an *edf* of 3 indicates an approximately cubic effect of a covariate on the outcome. The *edf* also gives an indication of how much penalization of a smooth function took place and thus may serve as a way to select the number of basis functions K (Wood, 2017, pp. 242–243). Having the *edf* close to K indicates a small penalization. Thus, it may be appropriate to increase K in such a case to describe the shape of the function.

Parameter Estimation

The main GAMM fitting routine in the `mgcv` package is the `gam()` function. The `bam()` function in the `mgcv` package provides an alternative for very large data sets such as eye-tracking data. Thus, we use the `bam()` function to fit the generalized additive logistic regression model. Below, we describe the details of the implementation in the `mgcv` package for the model.

The generalized additive logistic regression model can be written as a GLMM:

$$\text{logit}[P(y_{tlji} = 1 \mid \theta_j, \zeta_i)] = \mathbf{X}'\boldsymbol{\beta}, \tag{18}$$

where $\mathbf{X}$ is a design matrix having all components of the model including parametric components and all the basis functions ($b_{hk}(x_h)$ or $b_{hkk'}(x_{1h}, x_{2h})$ for $h = 1, \ldots, H$), $\boldsymbol{\beta}$ is a set of parameters including the coefficients of fixed effects ($\boldsymbol{\delta}$), random effects ($\mathbf{b} = [\boldsymbol{\theta}, \boldsymbol{\zeta}]'$), and the basis coefficients ($\boldsymbol{\gamma}$) (i.e., $\boldsymbol{\beta} = [\boldsymbol{\delta}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\gamma}]'$).

The "wiggliness" of smooth function f_h is controlled by a quadratic smoothing penalty. The multiple quadratic smoothing penalty for the model can be written as:

$$\sum_{h=1}^{H} \lambda_h \boldsymbol{\gamma}' \mathbf{S}_h \boldsymbol{\gamma}, \tag{19}$$

where λ_h is a smoothing parameter, $\boldsymbol{\gamma}$ is a vector of basis coefficients, and $\mathbf{S}_h$ is a penalty matrix embedded as a diagonal block in a matrix. For smooth functions, the elements of $\mathbf{S}_h$ are known and are determined by chosen basis functions ($b_{hk}(x_h)$ or $b_{hkk'}(x_{1h}, x_{2h'})$ (see Figure 1 for illustrations). For random effects, $\mathbf{S}$ is an identity matrix I (Kimeldorf & Wahba, 1970). The λ_h controls the trade-off between goodness of fit and smoothness of a smooth function. Having $\lambda_h \approx \infty$ results in a straight line estimate for f, whereas having $\lambda_h = \mathbf{0}$ leads to an unpenalized piecewise linear regression estimate.

Estimation has two components, (a) estimating model parameters ($\boldsymbol{\beta}$) and (b) estimating smoothing parameters (λ_h). In the `mgcv` package, the smooth parameter (λ_h) can be selected by either prediction error (GCV.Cp and GACV.Cp in the `mgcv` package) or marginal likelihood (REML and ML in the `mgcv` package). REML and ML are preferable to the other criteria, as they are less prone to local minima. In the `bam()` function of the `mgcv` package, fast restricted maximum likelihood estimation (fREML) can be used. However, fREML cannot be used to compare models with different fixed effects. Thus, in this study, ML was chosen in the `bam()` function.

Given smoothing parameters ($\boldsymbol{\lambda}$) and the variance matrix of the random effects $\mathbf{b}$ ($\Sigma = diag(\boldsymbol{\theta}, \boldsymbol{\zeta})$), parameters ($\boldsymbol{\beta}$) are estimated using a penalized iteratively reweighted least squares (PIRLS; Wood, 2017) with a default option, optimizer=c("outer," "newton"), in the `mgcv` package. In the PIRLS, the ML based $\boldsymbol{\lambda}$ can be inserted. Specifically, the following weighted least squares objective can be minimized to obtain $\hat{\boldsymbol{\beta}}$:

$$D(\boldsymbol{\beta}) + \phi \boldsymbol{\beta}' \Sigma^{-1} \boldsymbol{\beta} + \sum_{h=1}^{H} \lambda_h \boldsymbol{\beta}' \mathbf{S}_h \boldsymbol{\beta}, \tag{20}$$

where $D(\boldsymbol{\beta})$ is the model deviance ($D(\boldsymbol{\beta}) = 2\{l_{max} - l(\boldsymbol{\beta})\}$, where l is the log-likelihood) and ϕ is the scale parameter in an exponential family distribution (for binary responses, $\phi = 1$).

Imposing the penalty (Equation 19) is equivalent to having a prior on basis coefficients $\boldsymbol{\gamma}$ using a multivariate normal (*MVN*) distribution with mean vector $\mathbf{0}$ and the variance matrix $(\sum_h^H \hat{\lambda}_h \mathbf{S}_h)^{-1}$ (e.g., Wood, 2017):

$$\boldsymbol{\gamma} \sim MVN\left(\mathbf{0}, \left(\sum_{h=1}^{H} \hat{\lambda}_h \mathbf{S}_h\right)^{-1}\right), \tag{21}$$

where h is an index for a smooth function ($h = 1, \ldots, H$), $\hat{\lambda}_h$ is an estimated smoothing parameter for a smooth function f_h, and $\mathbf{S}_h$ is the known penalty matrix for smooth function f_h. The penalty matrix can be extracted using the `smoothCon` function in the `mgcv` package. To illustrate Equation 21, the same generated data used to explain the sum-to-zero constraint earlier was employed:

```
#extract penalty matrix S for a smooth
  #function of f1
#with the number of basis functions equal to 4
smooth.spec.object <- interpret.gam(y~s
  (x1,k=4))$smooth.spec[[1]]
SM<-smoothCon(smooth.spec.object,data=
  data,knots=NULL,absorb.cons=TRUE)[[1]]
S<-SM$S
>S
[[1]]
```

	[,1]	[,2]	[,3]
[1,]	3.039366e+01	−3.015506e + 00	−5.654629e-16
[2,]	-3.015506e + 00	6.670763e + 00	1.250891e-15
[3,]	-5.654629e-16	1.250891e-15	2.345651e-31

```
#extract estimated penalty parameter
model<-bam(y~1+s(x1,k=4),data=data,
  family=binomial,method="ML")
lambda<-model$sp
lambda
>lambda
   s(x1)
2587.993
#extract basis coefficient estimates
gamma<-model$coefficients
gamma
>gamma
```

(Intercept)	s(x1).1	s(x1).2	s(x1).3
5.311421e-01	−6.570583e-04	5.196194e-06	5.196194e-06

Basis coefficient estimates ($[-6.570583e-04, 5.196194e-06, 5.196194e-06]'$) are obtained with $MVN(\mathbf{0}, (2587.993S)^{-1})$, where $\lambda = 2587.993$ and S is the 3×3 penalty matrix shown above (instead of 4×4 due to the sum-to-zero constraint).

As mentioned earlier, the random effects (θ_j, ζ_i) are equivalent to a smooth with penalty matrix I (i.e., $\mathbf{S} = I$, where I is an identity matrix). That is, the variances of the random effects (θ_j and ζ_i) are the inverse of the estimated smoothing parameters ($\hat{\boldsymbol{\lambda}}$): $Var(\theta_j) = (\hat{\lambda}_\theta I)^{-1}$ and $Var(\zeta_i) = (\hat{\lambda}_\zeta I)^{-1}$ where $\hat{\lambda}_\theta$ and $\hat{\lambda}_\zeta$ are estimated smoothing parameters for θ_j and ζ_i, respectively. The `gam.vcomp` function in `mgcv` converts smoothing parameter estimates to the variance estimates of the random effects.

First Time Point Data Treatment for AR Effects

As illustrated in Table 1, there are missing values at the first time point ($t = 1$) for the three AR covariates ($y_{(t-1)lji}$ for the temporal AR; $d_{(t-1)lji}$, and $c_{(t-1)lji}$ for the time-varying spatial lag effects). In this study, the first time points were deleted prior to estimation with minimal consequences because of two reasons (as discussed in detail in Cho et al., 2018). First, for eye-tracking data, the number of time points is often larger than 100 (e.g., Mirman et al., 2011), so that the effect of deleting the first time point on parameter estimation is minimal (Hsiao, 2003). Second, the critical eye gaze data concerned fixations made between 180 ms and 1,499 ms in our motivating example. This time window is expected to capture the processing of the critical information. Signal-driven fixations are expected approximately 200 ms after onset. Thus, a baseline of 20 ms (180–200 ms) is included to facilitate calculation of the AR, indicating that the first 20 ms is not a data point to be modeled.

Model Selection and Evaluation

Prior to adding focal effects (e.g., experimental condition effects) to the model, we considered candidate models that have different spatial and temporal correlations to find the best-fitting model to explain correlations and variabilities in the data. The Akaike information criterion (AIC; Akaike, 1974) and the Bayesian information criterion (BIC; Schwarz, 1978) developed for GAMM were used for model comparisons. Instead of using the number of parameters in the penalty term of AIC and BIC, corrected AIC (Wood et al., 2016) and BIC (Wood, 2017) for GAMM use the effective degrees of freedom (*edf*) in the penalty term of AIC and BIC. The corrected AIC and BIC are specified as follows:

$$\text{CorrectedAIC} = -2l(\boldsymbol{\beta}) + (2 \times edf) \tag{22}$$

and

$$\text{BIC} = -2l(\boldsymbol{\beta}) + (log(N) \times edf), \tag{23}$$

where $l(\boldsymbol{\beta})$ is the log-likelihood and N is the total number of observations. The log-likelihood and *edf* for GAMM can be extracted using the function `logLik.gam` for a fitted model in the `mgcv` package. After fitting a model in the `mgcv` package, the corrected AIC and BIC can be calculated using the `AIC()` and `BIC()` functions in the `mgcv` R package.

To show explained variance when adding spatial-temporal effects, the percentage deviance explained (D_e) for each model is calculated as follows:

$$D_e = \left(\frac{D_{null} - D}{D_{null}}\right) \times 100, \tag{24}$$

where D_{null} is the deviance for a null model having a fixed intercept only and D is the deviance of the fitted model. Based on D_e for each model, the percentage explained variance for a spatial-temporal effect is calculated as

$$D_{e.1} - D_{e.2}, \tag{25}$$

where $D_{e.1}$ is the percentage deviance explained for a complex model including the spatial-temporal effect and $D_{e.2}$ is the percentage deviance explained for a simpler model omitting the spatial-temporal effect.

In addition, the Pearson residuals are calculated for model checking, calculated as

$$\epsilon_{tlji} = \frac{y_{tlji} - E(y_{tlji})}{\sqrt{Var(y_{tlji})}}, \tag{26}$$

where $E(y_{tlji}) = P(y_{tlji} = 1 \mid \widetilde{\theta}_j, \widetilde{\zeta}_i)$ and $Var(y_{tlji}) = P(y_{tlji} = 1 \mid \widetilde{\theta}_j, \widetilde{\zeta}_i)\{(1 - P(y_{tlji} = 1 \mid \widetilde{\theta}_j, \widetilde{\zeta}_i))\}$. When the model is correct, the Pearson residuals follow approximately a standard normal distribution. A standardized residual for one observation from a time point, a trial, a person, and an item can be far from normally distributed. However, observations with a standardized residual exceeding 1.96 in absolute value is worth a close look for misfit. Furthermore, the Somers' rank correlation between the binary data y_{tlji} and the model-based probabilities were calculated as a measure of the ordinal predictive power of the model. The `somers2()` function in the `hmisc R` package (Harrell, 2019) was used to calculate Somers' rank correlation.

Testing

To test whether or not the unpenalized parametric fixed effects ($\boldsymbol{\delta}$) are 0 ($H_0 : \boldsymbol{\delta} = \mathbf{0}$), Wald tests using the covariance matrix (which can be extracted for a fitted model using the `vcov` function) are conducted in the `mgcv` package. To test whether or not each smooth function $f_h(x)$ is needed in the model, the following null hypothesis can be tested for each h (Wood, 2017, pp. 305–306): H_0: $f_h(x) = \mathbf{0}$ for all x in the range of interest. Under H_0, the test statistic T_r follows a chi-square distribution ($T_r \sim \chi^2_r$) with degrees of freedom r (Wood, 2013). In the output from the `mgcv` package, the r is called *ref.df*, the reference degree of freedom used for hypothesis testing.

Smooths have credible intervals around them, which are obtained by taking the quantiles from the posterior distribution of the $f_h(x_h)$ (Marra & Wood, 2012). To obtain the posterior distribution of the $f_h(x_h)$, a large number of replicated parameters are simulated from a posterior distribution of parameters $\boldsymbol{\beta}$ using a multivariate normal (*MVN*) distribution:

$$\boldsymbol{\beta} \sim MVN(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}}_{\boldsymbol{\beta}}), \tag{27}$$

where $\hat{\boldsymbol{\beta}}$ is the vector of parameter estimates and $\hat{\mathbf{V}}_{\boldsymbol{\beta}}$ is the covariance matrix of parameter estimates. Based on replicated parameters, the predicted smooth functions can be calculated using the following equation:

$$\hat{\mathbf{f}}_h = X\hat{\boldsymbol{\beta}}, \tag{28}$$

where X is the design matrix of parameters $\boldsymbol{\beta}$. The design matrix X includes the basis functions for smooth functions $\mathbf{f}_h$ evaluated at each value of the covariates x_h, and includes columns of zeros corresponding to parameters unrelated to the smooth functions.

Using the model we fit to illustrate Equation 21, $\text{logit}[P(y = 1)] = f_1(x_1) = sin(x_1 + 2)$, the 95% credible interval for a smooth function $f_1(x_1)$ can be generated in `R` as follows:

```
library(ggplot2)
#fitted model
model <- bam(y ~ 1 + s(x1,k = 4),data=data,
  family=binomial,method="ML")
#extract a vector of parameter estimates
coef <- coef(model)
#extract a covariance matrix of parameter
  estimates
vcov <- as.matrix(vcov(model))
#replicate parameters from a posterior
  #distribution 1,000 times
beta <- rmvn(n = 1000,coef,vcov)
dim(beta) #1000 replications times 4
  #parameters (1 intercept + 3 basis coeff.)
#extract a design matrix X for the model
X <- predict(model,type="lpmatrix")
dim(X) #5000 observations times 4 parameters
  #(1 intercept + 3 basis coeff.)
#f1 has results of 5000 observations by 1000
  #replications.
Mean.fi <- rep(NA,1000)
f1 <- X %*% t(beta)
for (i in 1:1000){
Mean.f1[i] <- mean(f1[i,])
}
#calculate .025 and .975 quantiles of
  #predicted values for f1 across 1000
  #replications
CI <- apply(f1,1,quantile, probs = c(0.025,
  0.975))
lower <- as.matrix(CI[1,]) #values of 0.025
  #quantile
upper <- as.matrix(CI[2,]) #values of 0.975
  #quantile
#create a plot of credible interval
data1 <- data.frame(data$x1,mean.f1,lower,
  upper)
ggplot(data1, aes(data.x1, mean.f1))+
geom_ribbon(aes(ymin=data1$lower,ymax=
  data1$upper), linetype = 2, alpha = 0.1)
```

Illustration

In this section, we illustrate a generalized additive logistic regression model using an empirical data set from Trude et al. (2013). We provide `R` scripts to conduct all analyses in this section in Appendix C.

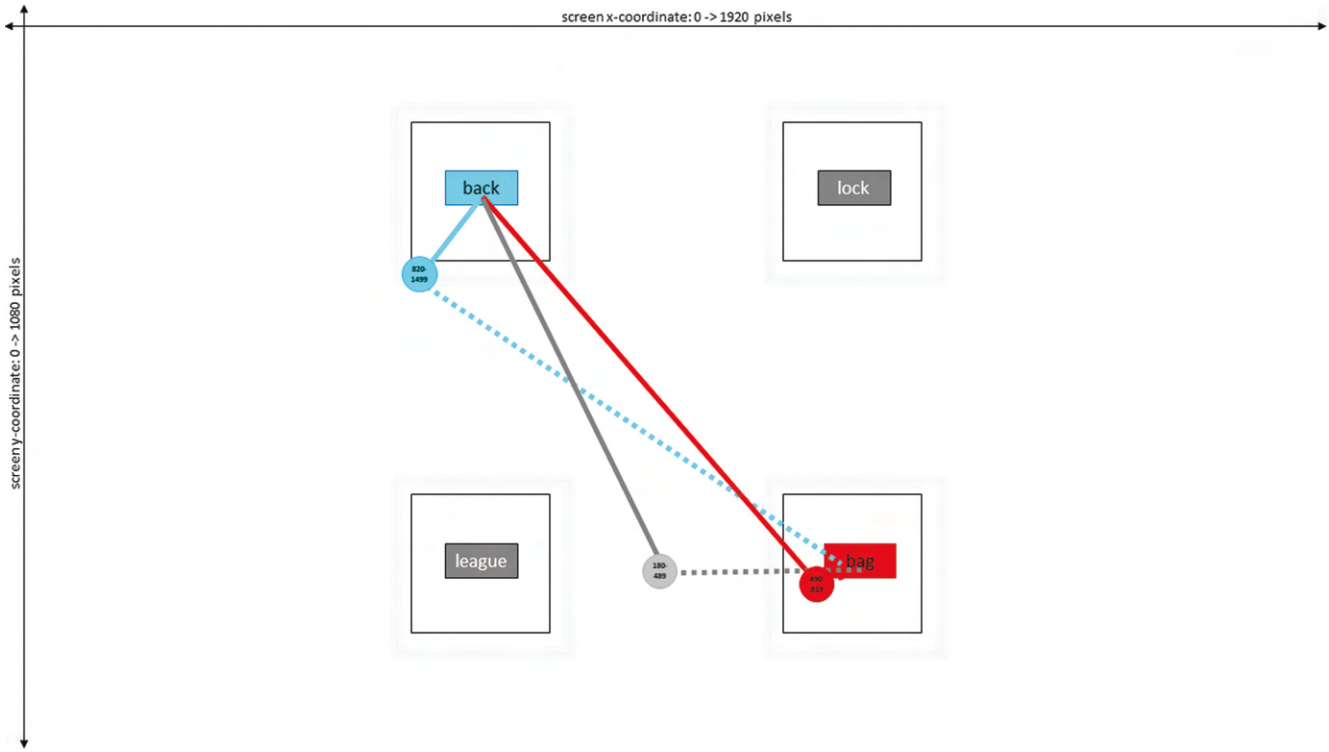
Eye-Tracking Experiment With Visual World Paradigm

A popular technique in the study of language processing (psycholinguistics) is the visual world eye-tracking paradigm (Tanenhaus et al., 1995). In this and related techniques, participants view a scene as they interpret or produce a spoken or signed utterance (e.g., Allopenna et al., 1998; Griffin & Bock, 2000; Lieberman et al., 2014; Pechmann, 1989). Eye movements are typically time-locked to an external event, such as the onset of a critical word in the utterance, and the researcher examines where the eyes are fixated over time as the person produces or interprets the utterance. This process is repeated over many ($\sim 10 - 400$) trials (e.g., utterances) per person to generate a sufficient quantity of data to make inferences regarding the cognitive processes involved.

The Eyelink eye-tracker (Eyelink, 2000; SR research) generated data with a temporal resolution of 1,000 Hz and a spatial resolution of .01° of visual angle. The data were downsampled in the temporal dimension to bins of 10 ms of time. As shown in Figure 2, on each trial, four pictures appeared on the screen in 200 × 200 pixel squares, then the participant heard an instruction to click one

Figure 2

Illustration of Spatial-Temporal Eye-Tracking Data



Note. Participants saw four images on each trial (for the purposes of illustration, the images have been replaced with their corresponding labels; participants viewed actual images). Images included the target referent (e.g., "back," represented by an image of a person's back), a competitor (e.g., "bag," represented by an image of a bag), and two unrelated images (e.g., an image of a person playing in a soccer "league," and an image of a "lock"). Participants were trained on image names prior to the eye-tracking task. The gray haloes around the images show the interest areas used to determine whether a person was fixating that image. The centroid of the target and competitor image areas correspond to the central point of each interest area. Superimposed on the image are circles corresponding to individual fixations to the target (blue), competitor (red), and unrelated objects (gray). The text in each circle indicates the time-range for each fixation in milliseconds, from 180 ms–1,499 ms following word onset. Solid lines indicate the distance between each fixation and the centroid of the target. Dotted lines indicate the distance between each fixation and the centroid of the competitor. This figure shows an example of a full screen display, with on-screen coordinates (in pixels) ranging from (0, 0) to (1,920, 1,080). See the online article for the color version of this figure.

of the four pictures, for example, "Click on back." The location of the target across the four possible picture locations was randomly varied for each trial. A training session preceded the trials so that participants were familiar with the image-label mappings. The critical eye gaze data concerned fixations made between 180 ms and 1,499 ms following the onset of the critical word (e.g., "back"). This window is expected to capture the processing of the critical word. The end of the time window corresponds to the point in time when, across all conditions, participants had identified the target, based on inspection of where the target fixation curves asymptoted. We use a measure of whether or not the participant was fixating the target at each of the 10-ms equally spaced time bins, resulting in 132 time points between 180 ms and 1,499 ms following the onset of the target word (e.g., "back"). In some cases, track loss due to difficulties in tracking eye-gaze resulted in data loss. In this dataset, 54% of the trials had 132 time points, and the remaining 46% had 47 to 131 time points.

The eye-tracking system automatically generated the fixation location based on the eye-tracker output. In the application, at each point in time, the on-screen fixation position was extracted in terms of x, y pixel coordinates relative to the 1920×1080 pixel sized computer screen. Note that on-screen fixations appeared within the pixel coordinates of the screen boundaries ([0, 0] to [1920, 1080]) and offscreen fixations occurred beyond those boundaries (including negative values for x, y coordinates and values > 1920 or > 1080 for x, y coordinates, respectively). Offscreen fixations are included in the analysis because these nontarget fixations also provide spatial information, with offscreen fixations varying considerably in the distance to the target and competitor. For further details about the processing and coding of the eye-tracking data, see Appendix D. If a fixation appeared within 50 pixels of the 200×200 pixel location of the target, this fixation was coded as a target fixation; otherwise it was coded as a nontarget fixation so that the data for the outcome variable are binary.

Experimental Design

Participants ($J = 60$) completed two experimental *sessions* spaced either 12- or 24-hr apart. Each participant was assigned to one of three between-subjects conditions that varied when those

sessions occurred (in the morning, a.m., or in the evening, p.m.). In two of the conditions, participants slept between sessions (p.m. to a.m.; p.m. to p.m.), and in the third condition participants did not sleep between sessions (a.m. to p.m.). Participants completed 512 trials in each of the two sessions. Of the 512 trials per session, half (256) were filler trials used to prevent the participants from guessing the target object, and were not submitted for analysis. The other half of the trials (256) were critical trials, each of which featured a target, a competitor that shared the same onset syllable as the target, and two other pictures. Eight item groups were created by manipulating the identity of the target and competitor, and each item group featured a different critical word that ended in the sound "-ag" (e.g., bag, rag, sag). Each item group had four items nested within it, resulting in 32 items in total. Each participant was exposed to each item four times with each of two talkers (a female and a male talker), resulting in the 256 critical trials (32 items $\times$ 4 repetitions $\times$ 2 talkers = 256). These 256 critical trials were combined with 256 filler trials in a random order. Participants completed these 512 trials (half of which featured the male talker, and half of which featured the female talker) twice—once in Session 1 and again in Session 2, resulting in 1,024 trials in total.

The data analyzed here represent a subset of the data described by Trude et al. (2013). In the experimental condition that we focus on, the eye-tracked participant heard a word like "back" and viewed a scene with four pictures: "back," "bag," "leak," and "league" (see Figure 2). For the present purpose of illustrating the space-time modeling of eye-tracking data using a generalized additive logistic regression model, we focus on the subset of trials in which participants heard a word like "back" in a scene containing pictures of the back (target), bag (competitor), leak, and league (fillers). The remaining critical trials featured other target types (e.g., bag, bake), and while these are of interest to the original empirical questions, they will not be addressed further here. Table 2 (top) illustrates the data structure for these 64 critical trials. Analysis of the full data set is reserved for a future article focusing on the substantive issues particular to the psycholinguistics questions in play with the same analysis approach. We measure fixations to the target, following the onset of the target word "back." The focus of the experiment is how listeners learn to accommodate an unfamiliar regional accent of English. Following prior work (Trude & Brown-Schmidt, 2012), we test processing of the speech of two *talkers*, a male and female. For the female talker, the words back and bag share a vowel and should be confusable; however, for the male talker, the words back and bag are pronounced with different vowels so they should be less confusable. Following prior work (Dahan et al., 2001), talkers with stereotypically male and female sounding voices are used so that listeners can easily distinguish the two talkers, allowing us to examine whether listeners can accommodate the speech of the accented talker. Thus, if the listener has learned this property of the male talker's accent, we expect more target ("back") and fewer competitor ("bag") fixations when perceiving the speech of the male talker. We measure interpretation using a binary measure of whether or not, at each (equally spaced) time point, the participant's gaze is fixated on the target.

In sum, the experimental conditions are summarized as follows:

- Each participant was assigned to one of three between-subjects manipulations that tested for effects of sleep on learning. Each participant completed two sessions that each occurred either in the morning (a.m.) or evening (p.m.), and that were separated by either 12 hr or 24 hr. These effects of sleep condition were coded using Helmert contrasts. The first contrast (*sleep1*) tested for an effect of

Table 2
An Experimental Design of the Empirical Study: Item Design (Top) and Person Design (Bottom)

		Session 1		Session 2	
Item group ID	Item ID	Talker female	Talker male	Talker female	Talker male
1	3	4	4	4	4
2	7	4	4	4	4
3	11	4	4	4	4
4	15	4	4	4	4
5	19	4	4	4	4
6	23	4	4	4	4
7	27	4	4	4	4
8	31	4	4	4	4
Total		32 Trials	32 Trials	32 Trials	32 Trials
Participant ID	Sleep	Talker Female	Talker Male	Talker Female	Talker Male
1	a.m. to p.m.	32 Trials	32 Trials	32 Trials	32 Trials
⋮	⋮	⋮	⋮	⋮	⋮
20	a.m. to p.m.	32 Trials	32 Trials	32 Trials	32 Trials
21	p.m. to a.m.	32 Trials	32 Trials	32 Trials	32 Trials
⋮	⋮	⋮	⋮	⋮	⋮
40	p.m. to a.m.	32 Trials	32 Trials	32 Trials	32 Trials
41	p.m. to p.m.	32 Trials	32 Trials	32 Trials	32 Trials
⋮	⋮	⋮	⋮	⋮	⋮
60	p.m. to p.m.	32 Trials	32 Trials	32 Trials	32 Trials

Note. ID indicates an identifier. Participants were randomly assigned to the three levels of *sleep* condition. Participant ID in table (bottom) indicates the sequence of IDs.

sleep, comparing the two groups who did sleep between sessions (p.m. to a.m. and p.m. to p.m. groups), with the group that did not sleep between sessions (a.m. to p.m. group). The second contrast (*sleep2*) tested whether the time of day at the second testing session affected performance, directly comparing the two groups who did sleep between sessions (p.m. to a.m. vs. p.m. to p.m. groups).

- *Session* (1, 2) was a within-subjects factor used to capture learning across the sessions.
- *Talker* (accented vs. unaccented) was a within-subjects factor to test for adaptation to the accented talker.

This between- and within-subjects design (or a mixed two-factor within-subjects factorial design; Keppel, 1991, Chapter 18) is summarized in Table 2 (bottom). In the design, the set of all possible experimental condition effects are as follows: (a) *sleep*, (b) *talker*, (c) *session*, (d) *sleep* × *talker*, (e) *sleep* × *session*, (f) *talker* × *session*, and (g) *sleep* × *talker* × *session*. As mentioned earlier, the analytic goal of the present study is to test these experimental condition effects, while accounting for all possible correlations and variabilities in the data, as captured with the generalized additive logistic model.

For each session, time points (which vary within trials) at Level 1 are nested within 64 trials at Level 2, which are cross-classified by 60 persons and eight items at Level 3. There were 32 items total in this experiment: four nested within each of eight item groups; in the present analysis we examine eight of the items. In this data structure, a trial is a series of time points which constitute time series. The total number of observations is 927,286.[4]

Research Question and Hypothesis

The focus of this research is whether learning a feature of a talker's regional accent, and then using this to shape online processing, is a cognitive process that is dependent on sleep. That is, we are interested in testing whether the Talker × Session interaction was stronger in the group who slept between sessions. The (male) accented talker spoke a dialect of American English in which the vowel in words like "bag" is shifted such that it sounds more similar to the vowel in a word like "bake." Prior research showed that, when exposed to this dialect, listeners readily learned this feature of the dialect and decreased their consideration of words like "bag" when hearing a word like "back," because the talker would have produced it with a different vowel (Dahan et al., 2008; Trude & Brown-Schmidt, 2012; Trude et al., 2014). Thus, we expect more target fixations for the accented talker compared with the unaccented talker. Furthermore, learning about the accented talker's accent is expected to increase across trials within a session, and between sessions. We predict that there will be more target fixations for the accented talker than for the unaccented talker, and that this talker difference will increase across sessions. If sleep plays an important role in this process, learning will be magnified in the groups that slept between sessions. If sleep plays a role in learning, then in the conditions where the participants slept, we expect a larger Talker × Session interaction.

Exploratory Descriptive Analysis: Spatial-Temporal Correlations

Trend and Temporal AR

For the exploratory descriptive analysis, the data were aggregated across persons and items, based on a binary response y_{tlji} for time point t, trial l, person j, and item i. The proportion of binary responses can be linearized by a logistic transformation in order to use statistics developed for linear models (Cox & Snell, 1989). For the empirical data, the proportion of binary fixations for *each* trial l over time can be calculated as $P_{tl} = \left(\sum_{j=1}^{J} y_{tlji}\right)/J$ and it bounds between 0 and 1. Thus, the proportion can be transformed using a logistic transformation as follows to calculate autocorrelations and partial autocorrelations for continuous outcomes:

$$log\frac{P_{tl}}{1-P_{tl}} = \text{logit}(P_{tl}). \tag{29}$$

Here, a subscript i is for a set of items which can be identified by a trial subscript and a person subscript. As a result, we did not consider summation across items. The logistic-transformed proportion measure of binary responses y_{tlji} is called *empirical logit*, which has the full range of real numbers.

Figure 3 presents box plots of trend, autocorrelations, and partial autocorrelations (all based on empirical logit), plotted using `R` to show variability in values across trials. Patterns in trend are explored by plotting the change in empirical logit over time, which results in the nonlinear (S-shaped) patterns observed in Figure 3. This S-shaped pattern may be due to the fact that at the beginning of the time-series participants do not know where the target is located, but by the end of the time series they have heard the target word and identified the target. Furthermore, the existence of trend and AR, and the order of AR are explored using the autocorrelations and partial autocorrelations based on the empirical logit. The autocorrelations for small lags presented in Figure 3 (middle) were large and positive, which may indicate that both AR and trend are needed to characterize the time series optimally. As shown in Figure 3 (bottom), the partial autocorrelations with order 1 (AR(1)) are clearly larger than 0, and those with a larger lag are nearly 0. This result indicates that only the AR(1) needs to be considered to model autocorrelation effects. Noticeably, there was little variability in partial autocorrelations across trials (SD = .004 for the two sessions). To summarize the descriptive investigation of the change process, a nonlinear trend and an AR(1) effect are suggested from this first and exploratory step.

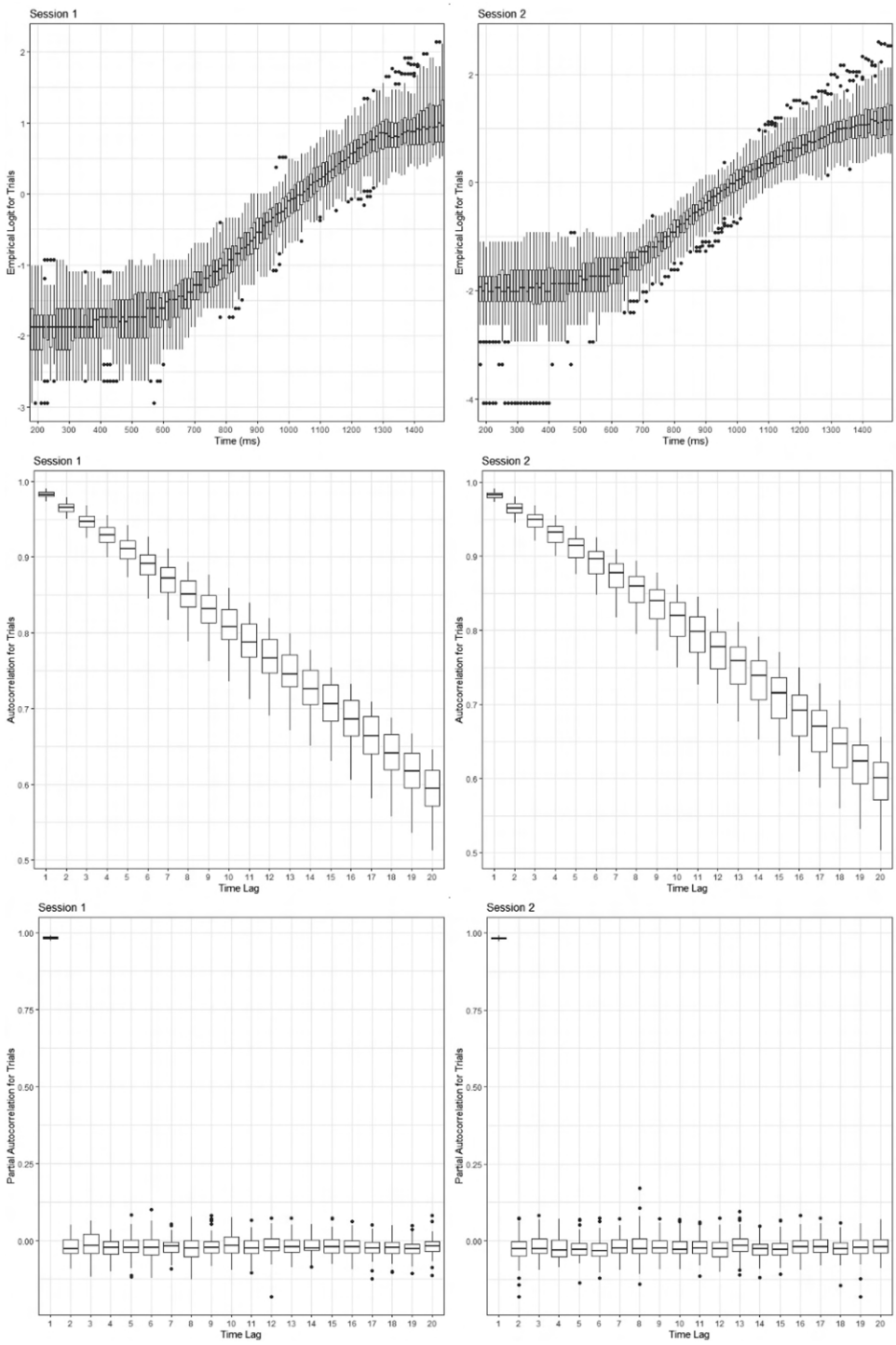
Spatial Correlations

In a similar way, we begin with an exploratory step for the spatial aspects of the data. Figure 4 (top) plots the semivariogram as a function of the (Euclidean) distance between fixation points on a computer screen. There are clear spatial correlations for distances between 0 and 2,000 pixels (with low semivariogram values indicating spatial dependence), which then level off for distances >

[4] Because the number of time points varies within trials, one cannot calculate the total number of observations by (132 time points × 64 trials × 2 sessions × 60 persons).

Figure 3

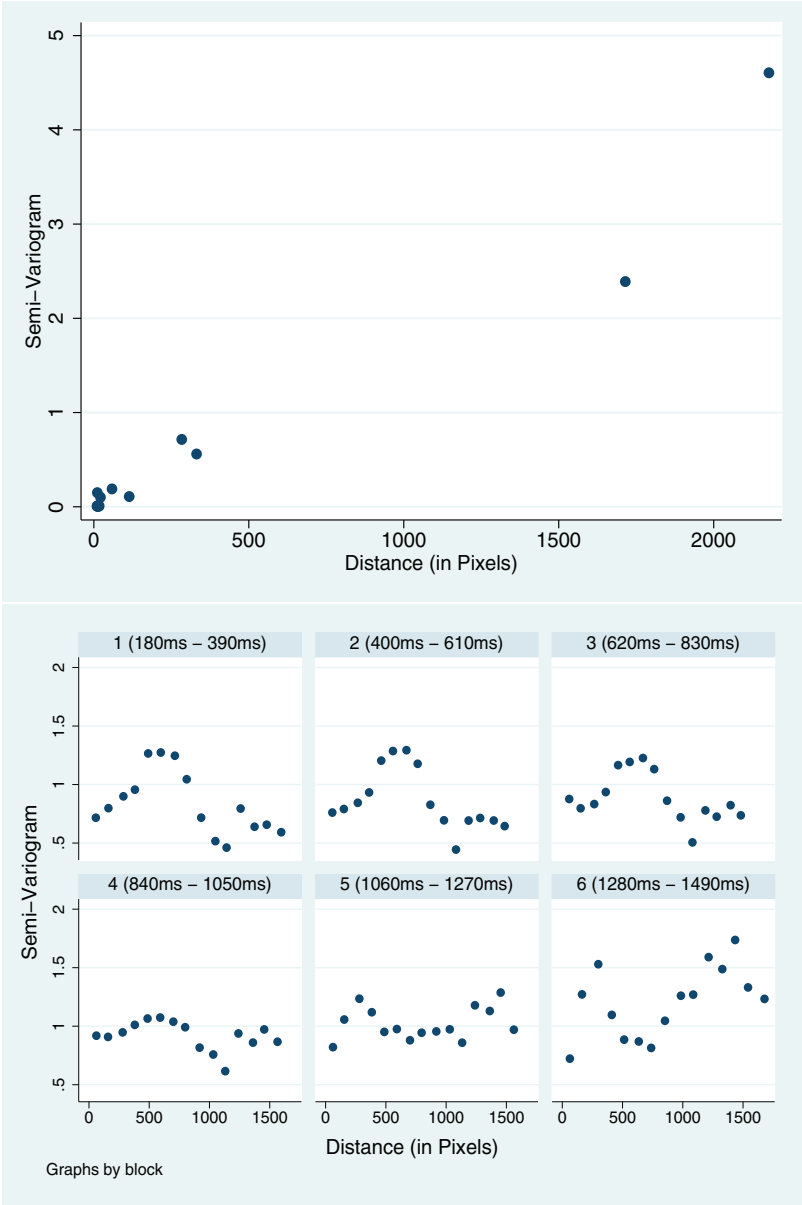
Patterns of Trends (Top), Autocorrelations (Middle), and Partial Autocorrelations (Bottom) Across Trials by Sessions



Note. Box plots are created to show variability across trials based on the empirical logit.

Figure 4

A Semivariogram of Residuals Across All Time Points (Top) and By Time Blocks (Bottom)



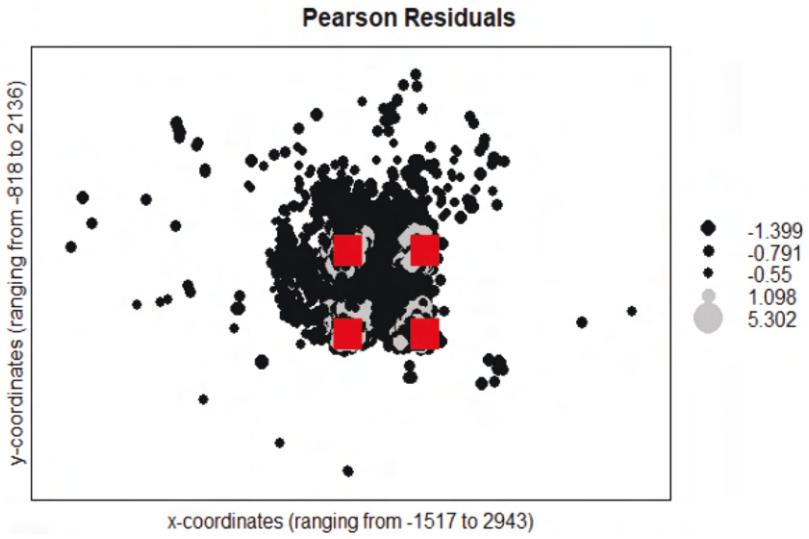
Note. A low semivariogam value indicates spatial dependence, while a large semivariogram value indicates spatial independence. See the online article for the color version of this figure.

1,000 (with high semivariogram values indicating spatial independence). The strength and patterns of spatial correlation were similar for the selected directions of north (0°), northeast (45°), east (90°), and southeast (135°) as four distinct directions, which indicates that isotropy can be assumed. Figure 5 presents a bubble plot of the residuals of the null model (Equation 1) at every screen position. The bubbles in Figure 5 represent the direction (gray bubbles for positive and black bubbles for negative) and magnitude (with larger bubbles indicating larger residuals) of the null model's residuals. High positive (large gray) bubbles were found in the positions of the four target images, which means that spatial correlations to be modeled exist for target fixations.

Spatial by Temporal Interactions

Figure 6 shows distributions of d_{tlji} (Euclidean distance between the fixation point and the centroid of the target location) and c_{tlji} (Euclidean distance between the fixation point and the centroid of the competitor location) across all observations (927,286). The values of d_{tlji} (shown in Figure 6, top) are clustered near zero reflecting coding of fixations on and near the target. Distances for

Figure 5
Pearson Residuals on Screen Locations



Note. The four rectangles in the plot indicate the locations of the four images. Positive residuals were plotted as gray bubbles and negative residuals were plotted as black bubbles; the size of the bubbles increase as the magnitude of the residuals increase. See the online article for the color version of this figure.

"nonfixation" exhibit two (or possibly three) clusters, corresponding to distances from the target centroid to the three other possible on-screen picture locations. The values of c_{tlji} (presented in Figure 6, bottom) exhibit two clusters corresponding to the distance between the target and the possible competitor locations (note that given the spacing of objects on the screen the Euclidean distances between target and competitor are the same for two possible competitor locations). Distances for "nonfixation" exhibit two clusters, corresponding to distances from the target centroid to the three other possible on-screen picture locations. For nonfixations, the cluster near zero reflects fixations on the competitor picture.

To explore whether spatial correlations differ over time, the semivariogram was calculated using the Pearson residuals of the null model (Equation 1) for six time blocks of 210 ms.[5] Figure 4 (bottom) presents semivariograms for the six time blocks. As shown in Figure 4 (bottom), spatial correlations differ by time blocks. The patterns in the semivariograms are similar across distances (in pixels) in the first three time blocks. Relatively high spatial dependence was observed in the fourth time block (between 840 ms and 1,050 ms), and decreased in the fifth and six time blocks (after 1,050 ms). Across the six blocks of time illustrated in the semivariograms, the dependencies are strongest in the middle time blocks. The relatively higher spatial dependencies in Blocks 4 and 5 may be due to the tendency to identify and look at the target in this time-period, followed by a tendency to look away from the target once the participant identified it, resulting in a drop in spatial dependency in the 6th time block.

To summarize descriptive and graphical analyses of the empirical data, there are spatial-temporal effects which must be accounted for in the generalized additive logistic model. The best-fitting model with spatial-temporal effects only will be selected prior to the addition of experimental condition covariates to the generalized additive logistic regression model.

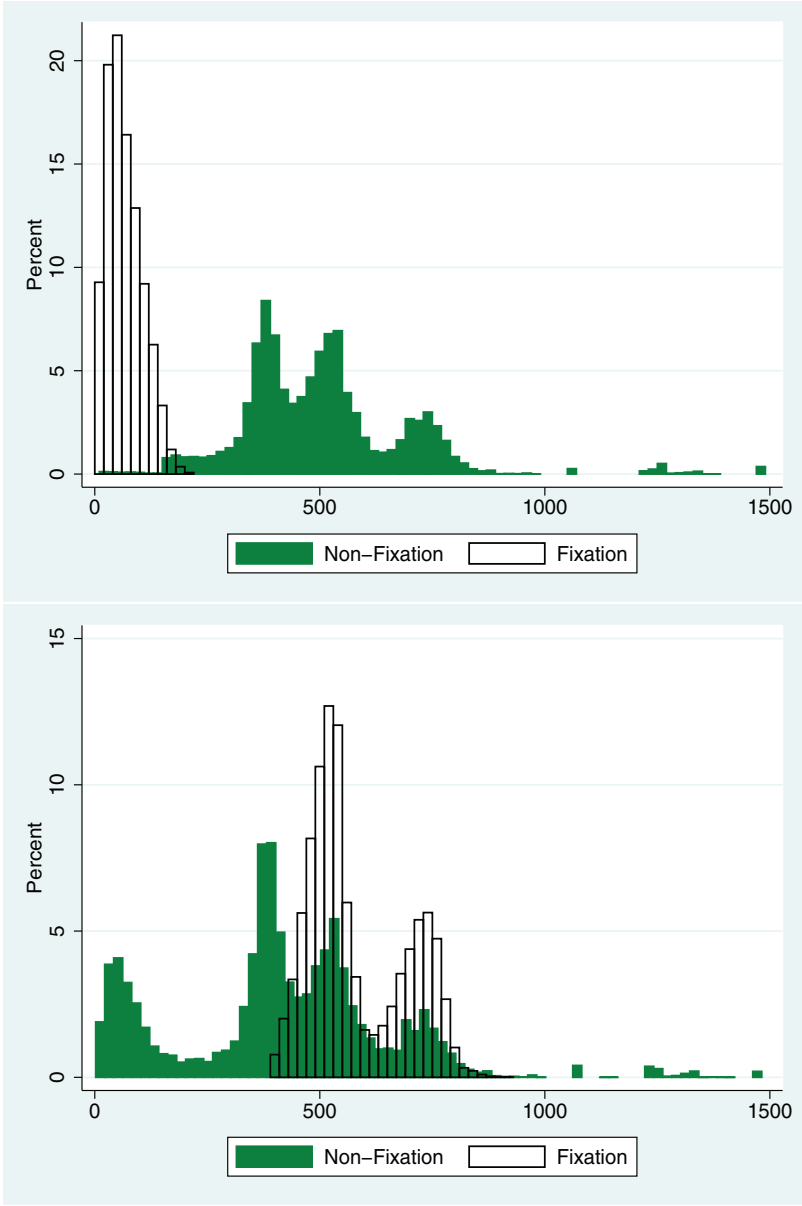
Analysis

A stepwise forward analysis strategy will be used to add temporal and spatial effects to the null model of Equation 1, as summarized in Table 3. Specifically, first, the smooth effect for time and the time-varying serial AR effect ($f_1(time_t)$ and $y'_{(t-1)lji}f_4(time_t)$, in Equation 17) will be added to the baseline model as Model A. Second, the smooth effect for trials and its interaction with the smooth effect for time ($f_2(trial_l)$ and $f_3(time_t, trial_l)$, in Equation 17) will be added to Model A yielding model B. When the trial effect by session ($session'f(trial_l)$) was added to model session differences in learning over trials by way of a preliminary analysis, the model fit did not improve. Thus, the main trial effect ($f_2(trial_l)$) was considered instead. Third, the two time-varying spatial distance effects ($d'_{(t-1)lji}f_5(time_t)$ and $c'_{(t-1)lji}f_6(time_t)$, in Equation 17) will be added to Model B yielding Model C. Finally, the spatial correlation effect ($f_7(xposition_{tlji}, yposition_{tlji})$ in Equation 17) will be added to Model C yielding Model D. To conclude, the null model, and Models A, B, C, and D will be considered for model selection regarding the spatial and temporal effects.

Before presenting the candidate models to account for the condition effects, we explain how the corresponding covariates are coded. Effect coding was used for *talker* (female = –.5; male = .5) and *session* (1 = –.5; 2 = .5) covariates. Effect coding was

[5] Block 1 = [180 ms, 390 ms], Block 2 = [400 ms, 610 ms], Block 3 = [620 ms, 830 ms], Block 4 = [840 ms, 1,050 ms], Block 5 = [1,060 ms, 1,270 ms], and Block 6 = [1,280 ms, 1,490 ms].

Figure 6
Distributions of Spatial Lag Covariates



Note. Distributions of the Euclidean distance (on x axis) between a fixation point and the centroid of a target location (top) and the Euclidean distance (on x axis) between a fixation point and the centroid of a competitor location (bottom) across all observations. See the online article for the color version of this figure.

considered instead of dummy coding to detect interaction effects of covariates (e.g., *talker* by *session*) instead of simple effects (e.g., *talker*'s effect within a *session*). Helmert coding was used for the three levels of the *sleep* covariate because we are interested in testing the two contrasts which compare each level to the subsequent levels: *sleep1* covariate (a.m. to p.m. = –.5; p.m. to a.m. = .25; p.m. to p.m. = .25) and *sleep2* covariate (a.m. to p.m. = 0; p.m. to a.m. = –.5; p.m. to p.m. = .5). In addition, in order to stabilize model fitting, unit distances were calculated prior to creating distance lag covariates (d_{tlji}/261.81 where 261.81 is the standard deviation across all observations for d_{tlji}; and c_{tlji}/221.68 where 221.68 is the standard deviation across all observations for c_{tlji}). Table 1 illustrates these distance covariates.

Based on a selected model regarding spatial-temporal effects (presented in Table 3), a model will be estimated with the following additional effects: all possible three-way interactions, two-way interactions, and main effects of *talker*, *session*, *sleep1*, and *sleep2*

Table 3
Model Selection for an Empirical Study

Added covariates to the preceding model	Log lik.	*edf*	Corrected AIC	BIC
Null (Equation 1)	−572468.73	66.95	1145071	1145857
Model A: $f_1(time_t)$ and $y'_{(t-1)lji} f_4(time_t)$	−45,378.19	79.34	90,915	91,847
Model B: $f_2(trial_l)$ and $f_3(time_t, trial_l)$	−45,366.99	86.56	90,907	91,923
Model C: $d'_{(t-1)lji} f_5(time_t)$ and $c'_{(t-1)lji} f_6(time_t)$	−45,203.52	98.24	90,604	91,757
Model D: $f_7(xposition_{tlji}, yposition_{tlji})$	—	—	—	—

Note. Log lik. indicates a log-likelihood value; Covariates were sequentially added to each model. For example, $f_1(time_t)$ and $y'_{(t-1)lji} f_3(time_t)$ were added to the null model (Equation 1) as Model A; There was a convergence problem with Model D. Thus, results were not presented in Table 3.

(*talker:session:sleep1*, *talker:session:sleep2*, *talker:sleep1*, *talker:sleep2*, *session:sleep1*, *session:sleep2*, *talker:session*, *talker*, *session*, *sleep1*, and *sleep2*) will be added to the selected model. In addition, the model without the spatial-temporal effects will also be fit to the same data set to show the consequences of ignoring spatial-temporal effects in detecting the experimental condition covariates.

Results

Table 3 presents the model fit information for the five candidate models with temporal and spatial effects. Results from the CRS are reported in Table 3. The TPRS results were similar. There was a convergence problem with Model D when the spatial correlation was considered. To explore why spatial smoothing over gaze position (*x*, *y* coordinates on the screen) did not work in Model D, we fit the generalized additive logistic regression model only with the spatial smooth function ($\text{logit}[P(y_{tlji} = 1)] = f_7(xposition_{tlji}, yposition_{tlji})$).[6] Figure 7 shows the result of this model, that is, the effect of *x*, *y* coordinates on the probability of target fixations.[7] The effect of the *x*, *y* coordinates on the probability of target fixations is 0 (.5 probability of a target fixation) when the *x*, *y* coordinates are in the four picture locations. As expected, the effect of the *x*, *y* coordinates on the probability of target fixations becomes negative when the gaze locations moves to the right-top side of the screen (as presented by the negative values of the contour lines and cold colors). In Figure 7, there is a contour line of 0 that encircles the four images. These results imply that while there is a spatial correlation (illustrated by the contour lines), its effect cannot differentiate between the four picture locations (i.e., fixations points within the four picture locations share the same effect of 0 [on the logit scale] on the target fixations). Among the converged models, Model C which includes all possible spatial-temporal effects except for the spatial correlation effect (listed in Table 3) had the smallest corrected AIC and BIC and the percentage explained deviance is 92.50%. The time-varying temporal AR effect substantially contributed to variance explained in the model (explained deviance = 87.22%), and all remaining effects contributed to a small amount of variance explained in the model (explained deviance = 5.28%).

Next, the main effects and the interaction effects for the experimental condition covariates are added to Model C, which is Model E as shown in Table 4. For Model E, about 12 mins (user time in `R`) were required to obtain results on a 2.81GHz computer with 16.0 GB of RAM. For all smooths in Model E, the *k*-index was close to 1 and the *edf* was lower than *K*, indicating the number of smooth functions ($K = 10$) is enough to obtain a good fit. In addition, 94.83% of observations (927,286) had a smaller than $|1.96|$ Pearson residual ($M = .000$, $SD = 1.008$). Finally, Somers' rank correlation between binary data and predicted probability (based on the results of Model E) was .994. These results indicate that Model E provides an adequate description of the data. We first interpret the results of controlling parameters (i.e., spatial-temporal effects and random variability) and then interpret the results of focal parameters (i.e., experimental condition effects).

Spatial-Temporal Effects and Random Variability

The significant and large fixed temporal AR(1) effect (EST = 8.867, $SE = .046$, p-value $<$ 2e-16)[8] suggests that there is a strong carry-over effect, due to the tendency for the eyes to be at a similar fixation position from one 10-ms time bin to the next. Regarding smooth terms, all smooths were statistically different from 0 (indicated by p-value $< .05$ at $\alpha = .05$). The patterns for each smooth term (i.e., partial effect, centered at 0) are presented in Figure 8 and Figure 9.

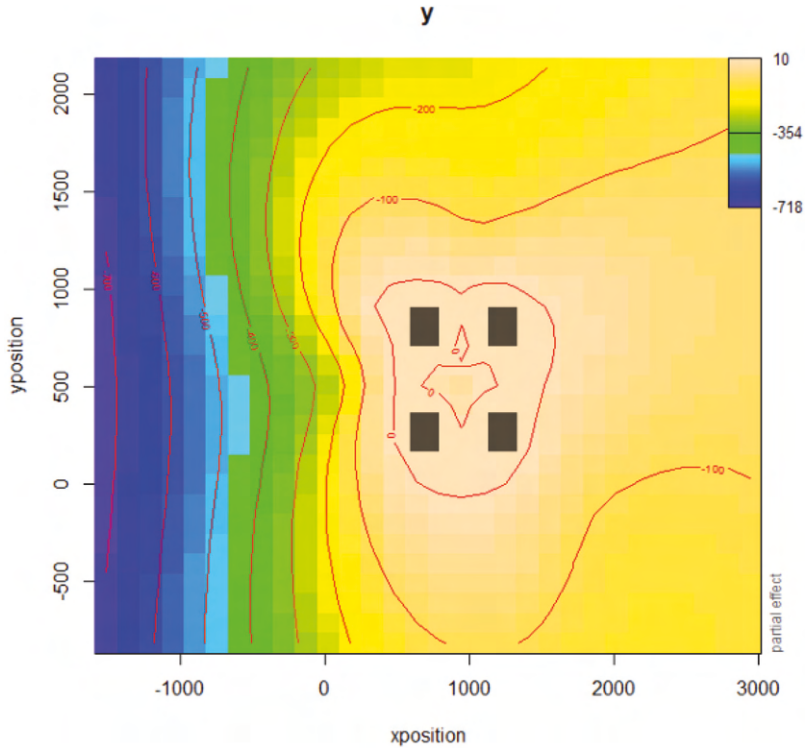
As shown in Figure 8 (top), smaller AR effects were observed in the middle range of time points, and larger AR effects were found at the beginning and end time points. This time varying AR likely results from the fact that between the beginning and end of the time-window of analysis the participants go from a state of not knowing what the target is to identifying and fixating on the target. At the beginning, we expect few target fixations (many consecutive 0's), and at the end we expect many target fixations (many consecutive 1's) whereas the middle is a period of change, resulting in a lower AR. Significant time-varying spatial lag effects for the target and the competitor were found. A time-varying spatial lag effect for the target (shown in Figure 8, middle) reflects the fact that the probability of a target fixation increased at *time t* when the fixation at *time t* − 1 was close to the target (controlling for the other effects). This time-varying spatial lag effect was attenuated over time, likely due to the fact that many fixations were

[6] The `mgcv` code is as follows: `bam(y ~ s(xposition; yposition; bs = "tp", k = 100), data = data, family = binomial, method = "ML")`. To check whether an alternative spatial smoother provides the same results, we also considered a Matérn based Gaussian process smoother. Similar patterns were found between the TPRS and Matérn based Gaussian process smoother.

[7] Figure 7 is a contour plot in which the effect of the *x*, *y* coordinates on target fixations is presented on the logit scale. Lines in the contour plot are drawn by connecting the *x*, *y* coordinates on which a logit value occurs.

[8] As shown in Appendix B, the time-varying AR(1) effect was modeled with an ordered factor. The ordered factor included the fixed AR(1) effect and the reference smooth.

Figure 7
*The Effect of x and y Coordinates on the Probability of Target Fixation (**y**)*



Note. Four rectangles in the figure indicate the locations of the four images. See the online article for the color version of this figure.

already on the target at later time points. This type of ceiling effect is common in eye-tracking data, as after a few hundred milliseconds most participants have interpreted the critical word (e.g., "back"), and have begun to fixate on the corresponding target picture in the visual display, in order to click on that image. As a result, at later time-points in the analysis window, there will be very high rates of target fixations, and thus a ceiling effect. In addition, the time-varying spatial lag effect for the competitor (shown in Figure 8, bottom) up to around time point 600 indicates that the probability of a target fixation at *time t* increased the farther away the eyes were from the competitor at *time* $t-1$. After time point 600, this effect attenuated, possibly because the competitor was only a weak pull on fixations once the target became more strongly activated. This effect likely results from the competition dynamics that guide lexical access, where upon hearing a word (e.g., "back"), multiple candidates initially compete for recognition (e.g., "back" vs. "bag"). Then, as the word unfolds in time, the competitor ("bag") becomes less active, as the target ("back") is consistent with the unfolding spoken word but the competitor is not (Allopenna et al., 1998).

Furthermore, Figure 9 (top) presents a result of $f_6(time_t, trial_l)$. Figure 9 (top) is a contour plot in which the effect of the interaction between time and trial on a target fixation is presented on the logit scale. Black lines in the contour plot are drawn connecting the (time, trial) coordinates where the same given value occurs. In the figure, the same effect of 0 in the middle (18th–48th trials) indicates that the trajectory (time effect) was not different across the middle range of trials; and the effect of $\pm.05$ at the beginning and end of trials shows that the trajectory (time effect) was different across trials (although the effect was very small). This interaction between trend and trial may reflect a learning effect in the early trials, and a fatigue process in the later trials, resulting in trajectories that were similar across the middle range of trials, and dissimilar for early and late trials. In addition, as shown in Figure 9 (middle), there was a S-curve trend effect over time, reflecting an initial increase in target fixations over the time-window of the time series that later drops off, likely due to participants locating and clicking on the target and then looking away (see Yoon & Brown-Schmidt, 2018 for a similar pattern). Furthermore, as presented in Figure 9 (bottom), there was little change in target fixations over the sequence of trials. Nonignorable variability was observed across persons and items, respectively (*SD* of random person effect = .508; *SD* of random item effect = .177, obtained from the `gam.vcomp` function).

Experimental Condition Effects

Regarding experimental condition effects, an interaction between *session* and *sleep2* (which tests for time-of-day effects) was significant (EST = $-.211$, *SE* = .053, *p*-value $< .0001$), controlling for the other effects. The significant interaction effect suggests that the group tested in the morning (p.m. to a.m.) at the second time showed an increase in target fixations from Session 1 to Session 2 (.455 → .500), as opposed to the group always tested in the evening (p.m. to p.m.; .526 → .518), which showed a decrease,

Table 4
Results for an Empirical Study

Parameters	Model E EST	Model E *SE*	Model E *z*-value	Model E *p*-value	Model F EST	Model F *SE*	Model F *z*-value	Model F *p*-value
Parametric coefficients								
δ_0[Intercept]	**−3.972**	0.104	−38.361	< 2e − 16	**−0.658**	0.099	−6.638	.000
δ_1[OFylag1]	**8.867**	0.046	193.578	< 2e − 16	—			
δ_2[talker]	**0.136**	0.022	6.268	.000	**0.126**	0.004	28.166	< 2e − 16
δ_3[session]	**0.074**	0.022	3.412	.001	0.003	0.004	0.559	.576
δ_4[sleep1]	−0.091	0.188	−0.481	.630	−0.090	0.226	−0.397	.691
δ_5[sleep2]	0.179	0.163	1.096	.273	0.198	0.196	1.012	.312
δ_6[talker:session]	**−0.087**	0.043	−1.993	.046	**−0.088**	0.009	−9.830	< 2e − 16
δ_7[talker:sleep1]	−0.038	0.061	−0.615	.538	**−0.117**	0.013	−9.275	< 2e − 16
δ_8[session:sleep1]	0.033	0.062	0.536	.592	**0.038**	0.013	3.021	.003
δ_9[talker:sleep2]	−0.045	0.053	−0.844	.398	**−0.043**	0.011	−3.882	.000
δ_{10}[session:sleep2]	**−0.211**	0.053	−3.963	.000	−0.234	0.011	−21.227	< 2e − 16
δ_{11}[talker:session:sleep1]	−0.099	0.123	−0.810	.418	0.001	0.025	0.021	.983
δ_{12}[talker:session:sleep2]	0.038	0.106	0.355	.722	**0.178**	0.022	8.073	.000

Parameters	Model E *edf*	Model E *Ref.df*	Model E Chi.sq	Model E *p*-value	Model F *edf*	Model F *Ref.df*	Model F Chi.sq	Model F *p*-value
Random effects								
θ_j[s(person)]*	55.434	57	1,645.49	< 2e − 16	56.95	57	66,715	< 2e − 16
ζ_i[s(itemid)]*	6.794	7	272.62	< 2e − 16	6.99	7	10,198	< 2e − 16
Smooth terms								
$f_1(time_t)$[s(time)]	6.791	7.668	453.30	< 2e − 16	—			
$f_2(trial_l)$[s(trialseq)]	2.273	2.831	14.96	.00,192	—			
$f_3(time_t, trial_l)$[ti(trialseq,time)]	3.347	4.350	12.56	.02,167	—			
$y'_{(t-1)lji} f_4(time_t)$[s(time):OFylag1]	6.285	7.377	283.22	< 2e − 16	—			
$d'_{(t-1)lji} f_5(time_t)$[s(time):tdlag1]	2.042	2.080	219.58	< 2e − 16	—			
$c'_{(t-1)lji} f_6(time_t)$[s(time):cdlag1]	6.278	7.221	86.49	1.27e-15	—			

Model E Log Lik.	Model E *edf*	Model E Corrected AIC	Model E BIC	Model F Log Lik.	Model F *edf*	Model F Corrected AIC	Model F BIC
−45,166.45	107.2,141	90,547.32	91,806.01	−571707.3	75.94,759	1,143,566	1144,458

Note. Model E is the selected model for result interpretations; Model F includes all effects of Model E, except for spatial and temporal effects; — indicates that an effect is not modeled; Values in bold indicates significance at the 5% level for parametric coefficients; **SD* of random person effect = .508; *SD* of random item effect = .177, obtained from `gam.vcomp` function.

possibly due to fatigue. Also, there was a significant main effect of *talker* (EST = .136, *SE* = .022, *p*-value < 2*e* − 16) and of *session* (EST = .074, *SE* = .022, *p*-value = .001), controlling for the other effects. The effect of *talker* is consistent with prior work (Trude & Brown-Schmidt, 2012) such that there are more target fixations with the accented talker (.146 greater; *exp*[.136] = 1.146 odds ratio), as the talker's accent rules out the competitor as a candidate referent. The main effect of session reflects a general learning effect, with faster identification of the target overall in the second session (.077 greater; or *exp*[.074] = 1.077 odds ratio). All the other effects except these three effects (interaction between *session* and *sleep2*, main *talker*, and main *session* effects) were not significant.

As explained, a model without spatial-temporal effects (called Model F) was also estimated. Although Model F was fit in the `mgcv` package, it is equivalent to a GLMM with crossed random effects (random person effect and random item effect). As shown in Table 4, Model E fits much better than Model F based on corrected AIC and BIC. As expected, the *SE*s of most experimental condition effects (all except the *sleep1* and *sleep2* covariates) in Model F were underestimated, compared with the *SE*s of the experimental condition effects in Model E. The underestimated SEs resulted in statistical significance for all two-way interaction effects and for the three-way interaction of *talker:session:sleep2* in Model F. In other words, the expected three-way interaction is significant based on Model F. This finding shows that whether or not the temporal and spatial correlations are accounted for in the model critically impacts the theoretical conclusions regarding the fixed effects.

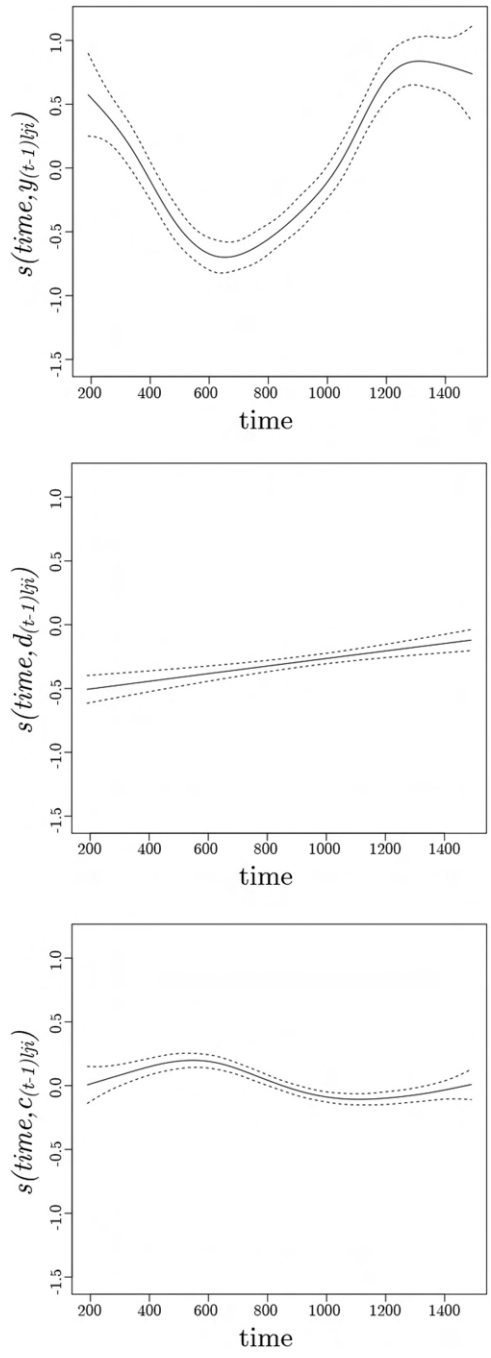
Simulation Study

Simulation Design and Analysis

A simulation study was designed to answer the following questions when the same conditions of the empirical study (i.e., the same number of time points, trials, persons, items, and sessions) and covariates are considered: (a) can the fixed experimental condition effects and their standard errors be recovered?; (b) what are the consequences of ignoring spatial and temporal effects in testing the experimental condition effects? The same true model (Model E in Table 4) was considered for the two questions. Model E was fit to the

Figure 8

Smooth Functions of the Time Varying AR (Top), the Time Varying Spatial Lag for Target (Middle), and the Time Varying Spatial Lag for Competitor (Bottom)



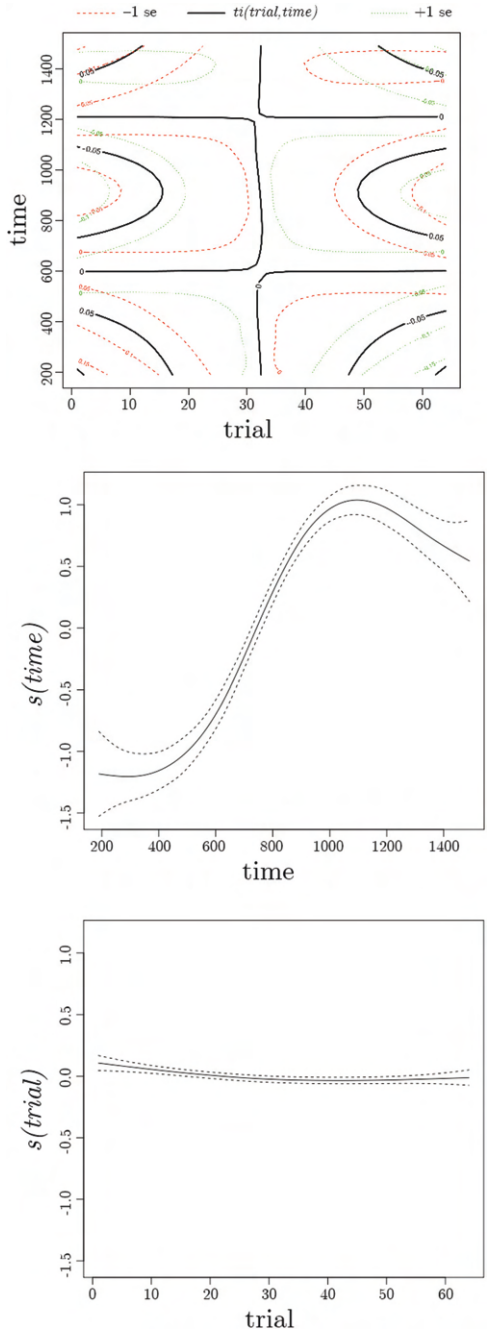
Note. The estimated effects are presented as solid lines and curves, with 95% credible intervals shown as dashed lines.

simulated data sets under Model E for question (a), and Model F was fit to the same simulated data sets for question (b).

The set of parameters for the data generation include the coefficients of the fixed effects ($\boldsymbol{\delta}$), values of random effects ($\mathbf{b} = [\boldsymbol{\theta}, \boldsymbol{\zeta}]'$), and the basis coefficients ($\boldsymbol{\gamma}$) (i.e., $[\boldsymbol{\delta}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\gamma}]'$). The estimates of $\boldsymbol{\beta}$ of Model E (presented in Table 4) for the empirical data were selected as true parameters. For the data generation, smooth functions in the true model (f_h, $h = 1, \ldots, 6$ in Equation 12) were constructed using

Figure 9

Smooth Functions of the Interaction Between Time and Trials Effect (Top), the Main Trend Effect (Middle), and the Main Trial Effect (Bottom)



Note. The estimated effects are presented as solid lines and curves, with 95% credible intervals shown as dashed lines. See the online article for the color version of this figure.

a prediction function in the `mgcv` package after fitting Model E to the empirical data: `predict(model1, type="iterms")`. Two-hundred replications were considered. As evaluation measures, bias was calculated to quantify the accuracy of parameter estimates, and root mean square error (RMSE) were calculated to quantify accuracy and parameter estimate variability. In addition, the mean standard error estimates (*M*(*SE*)) across 200 replications were compared with the standard deviations (*SD*) of the estimates to evaluate the accuracy of standard errors (*SE*s) for experimental condition effects (a subset of $\boldsymbol{\delta}$). The same estimation methods and the smooth functions in the `mgcv` package were used in the simulation study as in the empirical study.

Results

There were no convergence problems for the estimation of either Model E or Model F. Table 5 presents the results of parameter recovery (under Model E) and those of misspecification regarding spatial and temporal effects (under Model F). For the random effects (θ_j and ζ_i), the averaged results are shown across 60 persons for θ_j and across eight items for ζ_i. For the basis coefficients (γ) of smooth terms, averaged results are presented across the number of basis functions for each smooth term. For the parameter recovery under Model E (question (a)), bias for all estimates ($\hat{\boldsymbol{\beta}} = [\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\gamma}}]'$) was close to 0 (ranging from –.050 to .007). In addition, an acceptable level of RMSE for all estimates (ranging from .017 to .172) was obtained, compared with RMSE values in other simulation studies for GAMM (e.g., Bringmann et al., 2017). The M(*SE*) for parametric coefficients is close to the *SD* (the ratio range of *SD* to *M*(*SE*) = [.960, 1.111]), indicating that the estimated standard errors are approximately correct. Results from the estimation of Model F based on the simulated data were interpreted regarding the consequence of ignoring spatial and temporal effects in testing the experimental condition effects (question (b)). A larger bias and RMSE was found in Model F than in Model E. As expected, standard errors of all experimental condition effects except the two Helmert-coded *sleep* covariates (ratio range of *SD* to *M*(*SE*) = [4.115, 5.711]) were largely underestimated. To summarize, experimental condition effects and their standard errors were biased when spatial and temporal effects were not taken into account.

Summary and Discussion

Eye tracking is a widely used method to examine real-time cognitive processes in a variety of disciplines. However, the spatial-temporal correlations in binary coded eye-tracking data are rarely accounted for when testing effects of interest (e.g., experimental condition effects). This article was motivated by the need for statistical analysis methods for intensive binary spatial-temporal eye-tracking data. The benefits of the GAMM were illustrated by showing how to characterize the spatial-temporal correlations using descriptive statistics and graphics in R, how to estimate the generalized additive logistic regression model using the `mgcv` package, and how to interpret the results.

In the illustrative example, the results of the generalized logistic additive regression model showed that listeners were more likely to fixate on the target with the talker who spoke with a regional accent of American English in which the target and competitor words (e.g.,

Table 5
Results for a Simulation Study

Parameters	Under Model E				Under Model F			
	Bias	RMSE	*SD*	*M* (*SE*)	Bias	RMSE	*SD*	*M* (*SE*)
Parametric coefficients								
δ_0[Intercept]	−0.003	0.035	0.035	0.034	3.499	3.499	0.012	0.003
δ_1[OFylag1]	−0.001	0.019	0.019	0.018	—	—		
δ_2[talker]	0.002	0.017	0.017	0.018	0.027	0.038	0.026	0.004
δ_3[session]	0.001	0.027	0.027	0.026	−0.027	0.035	0.022	0.004
δ_4[sleep1]	0.004	0.023	0.023	0.023	−0.048	0.061	0.038	0.269
δ_5[sleep2]	−0.007	0.028	0.027	0.025	0.090	0.096	0.033	0.233
δ_6[talker:session]	0.000	0.034	0.034	0.036	−0.053	0.072	0.049	0.009
δ_7[talker:sleep1]	−0.001	0.050	0.050	0.050	−0.043	0.081	0.069	0.013
δ_8[session:sleep1]	−0.002	0.051	0.051	0.050	0.014	0.070	0.069	0.013
δ_9[talker:sleep2]	−0.002	0.047	0.048	0.044	−0.007	0.063	0.063	0.011
δ_{10}[session:sleep2]	−0.003	0.043	0.043	0.044	−0.051	0.078	0.059	0.011
δ_{11}[talker:session:sleep1]	−0.005	0.101	0.101	0.100	0.033	0.132	0.128	0.025
δ_{12}[talker:session:sleep2]	0.007	0.089	0.089	0.087	0.066	0.138	0.121	0.022
Random effects								
θ_j[s(person)]	0.001	0.033			0.000	0.211		
ζ_i[s(itemid)]	−0.003	0.028			0.000	0.041		
γ_k for smooth terms								
$f_1(time_t)$[s(time)]	−0.050	0.172			—	—		
$f_2(trial_l)$[s(trialseq)]	0.011	0.034			—	—		
$f_3(time_t, trial_l)$[ti(trialseq,time)]	0.001	0.022			—	—		
$y'_{(t-1)ljl} f_4(time_t)$[s(time):OFylag1]	0.000	0.032			—	—		
$d'_{(t-1)ljl} f_5(time_t)$[s(time):tdlag1]	0.000	0.068			—	—		
$c'_{(t-1)ljl} f_6(time_t)$[s(time):cdlag1]	0.000	0.024			—	—		

Note. Results under Model E are from the true model and those under Model F are from a misspecified model regarding spatial and temporal correlations; γ_k is a basis coefficient; — indicates that an effect is not modeled.

"back" and "bag") were pronounced with distinct vowels (as found in Dahan et al., 2008; Trude & Brown-Schmidt, 2012). Here, we also asked if this ability to learn and use information about the regional accent would change with sleep. The fact that the talker effect did not significantly differ between the groups that did versus did not sleep between sessions does not support this hypothesis. Relaxing the assumption of linearity in the GAMM enabled us to gain a better understanding of the data by capturing important structure that was overlooked in the model for which linearity was assumed. The following points from our illustration are worth noting for the controlling effects of the spatial and temporal correlations, in detecting experimental condition effects:

- Data-driven analyses uncovered an S-shaped trend in the probability of a target fixation over time within a trial. This S-shaped curve for target fixations is commonly observed and reflects a period of initial increasing consideration of the target, followed by looking away from the target (i.e., decreasing consideration of the target) after the participant has located the target.
- A strong temporal AR was observed to vary across time, with stronger AR at the beginning and end of the time-window of analysis. The AR was weaker in the middle of the time-window, reflecting stimulus-driven changes in fixation position.
- The time-varying spatial lag effects for target and competitor show how proximity to the target and distance from the competitor both increase the probability of a target fixation, particularly during initial processing of the target word (early in the time window). The fact that these effects were strongest in earlier time points likely reflects the initial period of ambiguity during which time the target and competitor are in competition.

With the simulation study, we showed that the accuracy of parameter estimates and their standard errors was satisfactory for the small sample numbers of persons and items similar to our illustrative data (132 time points, 64 trials, 60 persons, and eight items). Furthermore, ignoring the spatial and temporal effects led to biased inferences for the experimental condition effects of primary interest. In many applications where eye-gaze is used to understand cognitive processes, for example, in the study of speech perception, it is common to have more than 100 time points, 60 trials, and eight items (e.g., McMurray et al., 2002, 2009; Trude & Brown-Schmidt, 2012). For other types of research questions, it is common to have fewer than 60 persons in experimental studies, particularly when testing children or special populations (e.g., Creel, 2014; Trude et al., 2014). Additional research is needed to determine the number of persons under which the GAMM for binary data will perform optimally. While important, this question was not a focus in the present study.

Other Modeling Considerations

In this article, a generalized logistic additive regression model was illustrated using an empirical data set. The data set we selected has a data structure that is common for binary-coded eye-tracking data, that is, spatial-temporal correlations and random variability (random person effect and random item effect) in a multilevel (three-level) data structure. However, there are other potential effects that may be of interest, depending on the experiment design and spatial-temporal task structure. Below, we discuss other candidate approaches for modeling binary time series eye-tracking data.

Distance Between a Target and a Competitor

In the illustrative data set, we found that proximity to the target and distance from the competitor both increase the probability of a target fixation. By contrast, the distance between the target and competitor did not have a significant effect. This is likely because, given the arrangement of images within the experimental display (a grid of four images), the target and competitor were either separated by $\sim$ 300 pixels (when they were off-diagonal) or by $\sim$ 400 pixels (when they were on-diagonal). This difference may not have been substantive enough to affect target identification. While this distance effect was not a primary research interest, in other situations the primary research question may involve distance between target and competitor (e.g., Brown-Schmidt & Tanenhaus, 2008; Hanna & Brennan, 2007). For example, in a study where competitor fixations were a primary interest, Brown-Schmidt and Tanenhaus (2006) experimentally manipulated the distance between target and competitor and found that competitor fixations occurred earlier when this distance was smaller. In other cases, it is not only the distance between target and competitor that is relevant, but also whether there are potentially distracting images in between (see Ferreira et al., 2013), which could be modeled by weighting the distance measure.

Smooth Functions

Because the eye-tracking data in our illustration were from an experimental study, differences in smooths depend on between-subjects factors (e.g., three *sleep* conditions) and different smooth functions over trials by a within-subjects factor (e.g., two *sessions*) have also been considered in the generalized additive logistic model. However, these additional effects did not improve the model fit in terms of deviance or corrected AIC and BIC. However, we encourage researchers to consider such effects in the applications of the generalized additive logistic regression model for other data sets.

In addition to the time-varying spatial (distance) lag effects we illustrated, spatial smoothing over gaze position (x, y coordinates on the screen) can be used to model spatial correlations in eye-tracking data as a covariate in the generalized additive logistic regression model. As reviewed earlier, this kind of spatial smoothing was applied to the continuous time-series pupilometric data (van Rij et al., 2019). In addition, the spatial smoothing was considered in a logistic regression model to model spatial variances in disease risk for cross-sectional data (Siangphoe & Wheeler, 2015) and in a linear regression model to model the locations of voxels in brain activity level measurement (Wood, 2017, 7.2). In our illustration, there were convergence problems when the model had the spatial smooths over gaze positions (Model D in Table 3). A possible reason is that there is not enough information left in the data to estimate the spatial correlation after all other effects are accounted for. When there are significant x, y coordinate effects on the probability of target fixations in the other applications, it is expected that the spatial smoothing is an interesting GAMM application to eye-tracking data.

Conclusion

To conclude, this article shows that the GAMM approach is a feasible technique for modeling nonlinear effects of space-time information

in binary intensive time series eye-tracking data. In the current study, we present how a GAMM approach can be used to explore and test not only experimental condition effects, but also novel investigations regarding nonlinear spatial-temporal effects that influence visual attention and the underlying cognitive processes that guide the eyes. Furthermore, we show that accounting for spatial and temporal dynamics are critical in order to accurately test the effects of interest (e.g., experimental condition effects). The use of eye-tracking data to study cognitive processes is common across a variety of disciplines, yet data analytic techniques for analyzing these complex data are currently limited in practice. The GAMM approach is likely to be of increasing interest to researchers as eye-tracking techniques are extended to increasingly ecologically valid and complex real-world tasks.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558

Baayen, R. H. (2010). The directed compound graph of English. An exploration of lexical connectivity and its processing consequences. In E. Olson (Ed.), *New impulses in word-formation* (pp. 383–402). Buske.

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

Baayen, R. H., van Rij, J., de Cat, C., & Wood, S. N. (2018). Autocorrelated errors in experimental data in the language sciences: Some solutions offered by generalized additive mixed models. In D. Speelman, K. Heylen, & D. Geeraerts (Eds.), *Mixed-effects regression models in linguistics* (pp. 49–69). Springer International Publishing AG. https://doi.org/10.1007/978-3-319-69830-4_4

Baayen, R. H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language*, *94*, 206–234. https://doi.org/10.1016/j.jml.2016.11.006

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. https://doi.org/10.1016/j.jml.2007.09.002

Barr, D. J., Jackson, L., & Phillips, I. (2014). Using a voice to put a name to a face: The psycholinguistics of proper name comprehension. *Journal of Experimental Psychology: General*, *143*(1), 404–413. https://doi.org/10.1037/a0031813

Bartolucci, F., & Nigro, V. (2010). A dynamic model for binary panel data with unobserved heterogeneity admitting a $\sqrt{n}$-consistent conditional estimator. *Econometrica*, *78*(2), 719–733. https://doi.org/10.3982/ECTA7531

Box, G. E. P., & Jenkins, G. (1976). *Time series analysis: Forecasting and control*. Holden-Day.

Bringmann, L. F., Ferrer, E., Hamaker, E. L., Borsboom, D., & Tuerlinckx, F. (2018). Modeling nonstationary emotion dynamics in dyads using a time-varying vector-autoregressive model. *Multivariate Behavioral Research*, *53*(3), 293–314. https://doi.org/10.1080/00273171.2018.1439722

Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods*, *22*(3), 409–425. https://doi.org/10.1037/met0000085

Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language*, *54*(4), 592–609. https://doi.org/10.1016/j.jml.2005.12.008

Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science*, *32*(4), 643–684. https://doi.org/10.1080/03640210802066816

Chatfield, C. (2004). *The analysis of time series: An introduction* (6th ed.). Chapman; Hall/CRC. https://doi.org/10.1007/978-1-4899-2921-1

Cho, S.-J., Brown-Schmidt, S., De Boeck, P., & Shen, J. (2020). Modeling intensive polytomous time series eye tracking data: A dynamic tree-based item response model. *Psychometrika*, *85*(1), 154–184. https://doi.org/10.1007/s11336-020-09694-6

Cho, S.-J., Brown-Schmidt, S., & Lee, W.-y. (2018). Autoregressive generalized linear mixed effect models with crossed random effects: An application to intensive binary time series eye-tracking data. *Psychometrika*, *83*(3), 751–771. https://doi.org/10.1007/s11336-018-9604-2

Cho, S.-J., Brown-Schmidt, S., Naveiras, M., & De Boeck, P. (2020). A dynamic generalized mixed effect model for intensive binary temporal-spatio data from an eye tracking technique. In H. Jao & R. W. Lissitz (Eds.), *Innovative psychometric modeling and methods* (pp. 45–68). Information Age Publishing.

Cox, D. R., & Snell, E. J. (1989). *Analysis of binary data* (2nd Ed.). Chapman and Hall/CRC.

Craigmile, P. F., Peruggia, M., & Van Zandt, T. (2010). Detrending response time series. In S.-M. Chow, E. Ferrer, & F. Hsieh (Eds.), *Statistical methods for modeling human dynamics: An interdisciplinary dialogue* (pp. 213–240). Taylor & Francis.

Creel, S. C. (2014). Preschoolers' flexible use of talker information during word learning. *Journal of Memory and Language*, *73*, 81–98. https://doi.org/10.1016/j.jml.2014.03.001

Cressie, N. (1993). *Statistics for spatial data*. Wiley. https://doi.org/10.1002/9781119115151

Cressie, N., & Wikle, C. K. (2011). *Statistics for spatio-temporal data*. Wiley.

Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition*, *108*(3), 710–718. https://doi.org/10.1016/j.cognition.2008.06.003

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*(5-6), 507–534. https://doi.org/10.1080/01690960143000074

De Boeck, P. (2008). Random item IRT models. *Psychometrika*, *73*(4), 533–559. https://doi.org/10.1007/s11336-008-9092-x

de Haan-Rietdijk, S., Gottman, J. M., Bergeman, C. S., & Hamaker, E. L. (2016). Get over it! A multilevel threshold autoregressive model for state-dependent affect regulation. *Psychometrika*, *81*(1), 217–241. https://doi.org/10.1007/s11336-014-9417-x

Fahrmeir, L., & Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors. *Journal of the Royal Statistical Society Series C Applied Statistics*, *50*(2), 201–220. https://doi.org/10.1111/1467-9876.00229

Fang, X., & Chan, K.-S. (2015). Generalized additive models with spatio-temporal data. *Environmental and Ecological Statistics*, *22*(1), 61–86. https://doi.org/10.1007/s10651-014-0283-6

Ferreira, F., Foucart, A., & Engelhardt, P. E. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *Journal of Memory and Language*, *69*(3), 165–182. https://doi.org/10.1016/j.jml.2013.06.001

Finch, W. H., & Finch, M. H. (2018). A simulation study evaluating the generalized additive model for assessing intervention effects with small samples. *Journal of Experimental Education*, *86*(4), 652–670. https://doi.org/10.1080/00220973.2017.1339010

Fokianos, K., & Kedem, B. (2003). Regression theory for categorical time series. *Statistical Science*, *18*(3), 357–376. https://doi.org/10.1214/ss/1076102425

Eyelink. (2000). [Apparatus and software]. *SR Research*.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*(4), 274–279. https://doi.org/10.1111/1467-9280.00255

Hanna, J. E., & Brennan, S. E. (2007). Speakers eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, *57*(4), 596–615. https://doi.org/10.1016/j.jml.2007.01.008

Harrell, F. E. (2019). *Harrell Miscellaneous* (R package version 4.2-0). https://cran.r-project.org/web/packages/Hmisc

Hastie, T., & Tibshirani, R. (1990). *Generalized additive models*. Chapman & Hall.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*(1), 6–6. https://doi.org/10.1167/3.1.6

Hsiao, C. (2003). *Analysis of panel data*. Cambridge University Press. https://doi.org/10.1017/CBO9781139839327

Hung, Y., Zarnitsyna, V., Zhang, Y., Zhu, C., & Wu, C. J. (2008). Binary time series modeling with application to adhesion frequency experiments. *Journal of the American Statistical Association*, *103*(483), 1248–1259. https://doi.org/10.1198/016214508000000508

Irwin, D. E. (2004). Fixation location and fixation duration as indices of cognitive processing. *The Interface of Language, Vision, and Action: Eye Movements and the Visual World*, *217*, 105–133.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Kimeldorf, G. S., & Wahba, G. (1970). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *The Annals of Mathematical Statistics*, *41*(2), 495–502. https://doi.org/10.1214/aoms/1177697089

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*(4), 963–974. https://doi.org/10.2307/2529876

Lieberman, A. M., Borovsky, A., Hatrak, M., & Mayberry, R. I. (2014). Real-time processing of ASL signs: Effects of linguistic experience and proficiency. In W. Orman & M. J. Valleau (Eds.), *Proceedings of the 38th Boston University Conference on Language Development* (pp. 279–291). Cascadilla Press.

Lin, X., & Zhang, D. (1999). Inference in generalized additive mixed models using smoothing splines. *Journal of the Royal Statistical Society B*, *61*(2), 381–400. https://doi.org/10.1111/1467-9868.00183

Loo, K., van Rij, J., Jarvikivi, J., & Baayen, R. H. (2016). Individual differences in pupil dilation during naming task. In A. Papafragou, D. Grodner, D. Mirman, & J. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 550–555). Cognitive Science Society.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703. https://doi.org/10.1037/0033-295X.101.4.676

Marra, G., & Wood, S. N. (2012). On *p*-values for smooth components of an extended generalized additive model. *Scandinavian Journal of Statistics*, *39*(1), 53–74. https://doi.org/10.1111/j.1467-9469.2011.00760.x

McCamy, M. B., Otero-Millan, J., Di Stasi, L. L., Macknik, S. L., & Martinez-Conde, S. (2014). Highly informative natural scene regions increase microsaccade production during visual scanning. *Journal of Neuroscience*, *34*(8), 2956–2966. https://doi.org/10.1523/JNEUROSCI.4448-13.2014

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33–B42. https://doi.org/10.1016/S0010-0277(02)00157-9

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from "lexical" garden-paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language*, *60*(1), 65–91. https://doi.org/10.1016/j.jml.2008.07.002

Mirman, D., Yee, E., Blumstein, S. E., & Magnuson, J. S. (2011). Theories of spoken word recognition deficits in aphasia: Evidence from eye-tracking and computational modeling. *Brain and Language*, *117*(2), 53–68. https://doi.org/10.1016/j.bandl.2011.01.004

Nixon, J. S., van Rij, J., Mok, P., Baayen, R. H., & Chen, Y. (2016). The temporal dynamics of perceptual uncertainty: Eye movement evidence from Cantonese segment and tone perception. *Journal of Memory and Language*, *90*, 103–125. https://doi.org/10.1016/j.jml.2016.03.005

Oakes, L. M. (2012). Advances in eye tracking in infancy research. *Infancy*, *17*(1), 1–8. https://doi.org/10.1111/j.1532-7078.2011.00101.x

Oakes, L. M., Baumgartner, H. A., Barrett, F. S., Messenger, I. M., & Luck, S. J. (2013). Developmental changes in visual short-term memory in infancy: Evidence from eye-tracking. *Frontiers in Psychology*, *4*(OCT), 697. https://doi.org/10.3389/fpsyg.2013.00697

Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2011). A hierarchical latent stochastic differential equation model for affective dynamics. *Psychological Methods*, *16*(4), 468–490. https://doi.org/10.1037/a0024375

Pebesma, E. J. (2004). Multivariable geostatistics in S: The gstat package. *Computers & Geosciences*, *30*(7), 683–691. https://doi.org/10.1016/j.cageo.2004.03.012

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*, *27*(1), 89–110. https://doi.org/10.1515/ling.1989.27.1.89

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*(4), 413–425. https://doi.org/10.1016/j.jml.2008.02.002

R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Schuurman, N. K., Ferrer, E., de Boer-Sonnenschein, M., & Hamaker, E. L. (2016). How to compare cross-lagged associations in a multilevel autoregressive model. *Psychological Methods*, *21*(2), 206–221. https://doi.org/10.1037/met0000062

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Shadish, W. R., Zuur, A. F., & Sullivan, K. J. (2014). Using generalized additive (mixed) models to analyze single case designs. *Journal of School Psychology*, *52*(2), 149–178.

Siangphoe, U., & Wheeler, D. C. (2015). Evaluation of the performance of smoothing functions in generalized additive models for spatial variation in disease. *Cancer Informatics*, *14*(Suppl. 2), 107–116. https://doi.org/10.4137/CIN.S17300

Sullivan, K. J., Shadish, W. R., & Steiner, P. M. (2015). An introduction to modeling longitudinal data with generalized additive models: Applications to single-case designs. *Psychological Methods*, *20*(1), 26–42.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863

Thomas, L. E., & Lleras, A. (2009). Covert shifts of attention function as an implicit aid to insight. *Cognition*, *111*(2), 168–174. https://doi.org/10.1016/j.cognition.2009.01.005

Tremblay, A., & Newman, A. J. (2015). Modeling nonlinear relationships in ERP data using mixed-effects regression with R examples. *Psychophysiology*, *52*(1), 124–139. https://doi.org/10.1111/psyp.12299

Trude, A. M., & Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes*, *27*(7–8), 979–1001. https://doi.org/10.1080/01690965.2011.597153

Trude, A. M., Brown-Schmidt, S., & Fenn, K. M. (2013). *Contributions of sleep and exposure to talker-specific adaptation* [Poster presentation]. Annual meeting of the Psychonomic Society, Toronto, Canada.

Trude, A. M., Duff, M. C., & Brown-Schmidt, S. (2014). Talker-specific learning in amnesia: Insight into mechanisms of adaptive speech perception. *Cortex*, *54*, 117–123. https://doi.org/10.1016/j.cortex.2014.01.015

van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the time course of pupillometric data. *Trends in Hearing*, *23*, 1–22. https://doi.org/10.1177/2331216519832483

Viviani, P., & Swensson, R. G. (1982). Saccadic eye movements to peripherally discriminated visual targets. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(1), 113–126. https://doi.org/10.1037/0096-1523.8.1.113

Vogelzang, M., Hendriks, P., & van Rijn, H. (2016). Pupillary responses reflect ambiguity resolution in pronoun processing. *Language, Cognition and Neuroscience*, *31*(7), 876–885. https://doi.org/10.1080/23273798.2016.1155718

Warren, T., White, S. J., & Reichle, E. D. (2009). Investigating the causes of wrap-up effects: Evidence from eye movements and E-Z reader. *Cognition*, *111*(1), 132–137. https://doi.org/10.1016/j.cognition.2008.12.011

Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists*. Wiley.

Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., Bröker, F., Thiele, S., Wood, S. N., & Baayen, R. H. (2016). Investigating dialectal differences using articulography. *Journal of Phonetics*, *59*, 122–143. https://doi.org/10.1016/j.wocn.2016.09.004

Wood, S. N. (2013). A simple test for random effects in regression models. *Biometrika*, *100*(4), 1005–1010. https://doi.org/10.1093/biomet/ast038

Wood, S. N. (2017). *Generalized additive models: An introduction with R. Second edition*. Chapman and Hall/CRC.

Wood, S. N. (2019). *Mixed gam computation vehicle with automatic smoothness estimation* (published on the Comprehensive R Archive Network, CRAN). https://cran.r-project.org/web/packages/mgcv

Wooding, D. S. (2002). Eye movements of large populations: II. Deriving regions of interest, coverage, and similarity using fixation maps. *Behavior Research Methods, Instruments, & Computers*, *34*(4), 518–528. https://doi.org/10.3758/BF03195481

Wu, C. C., Kwon, O. S., & Kowler, E. (2010). Fitts's Law and speed/accuracy trade-offs during sequences of saccades: Implications for strategies of saccadic planning. *Vision Research*, *50*(21), 2142–2157. https://doi.org/10.1016/j.visres.2010.08.008

Yamashiro, A., Shrout, P., & Vouloumanos, A. (2019). Using spline models to analyze event-based changes in eye tracking data. *Journal of Cognition and Development*, *20*(3), 299–313. https://doi.org/10.1080/15248372.2019.1583231

Yarbus, A. L. (1967). Eye movements during perception of complex objects. In A. L. Yarbus (Ed.), *Eye movements and vision* (pp. 171–211). Springer. https://doi.org/10.1007/978-1-4899-5379-7_8

Yoon, S. O., & Brown-Schmidt, S. (2018). The influence of the historical discourse record on language processing in dialogue. *Discourse Processes*, *55*(1), 31–46. https://doi.org/10.1080/0163853X.2016.1193429

Zeger, S. L., & Qaqish, B. (1988). Markov regression models for time series: A quasi-likelihood approach. *Biometrics*, *44*(4), 1019–1031. https://doi.org/10.2307/2531732

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. Springer.

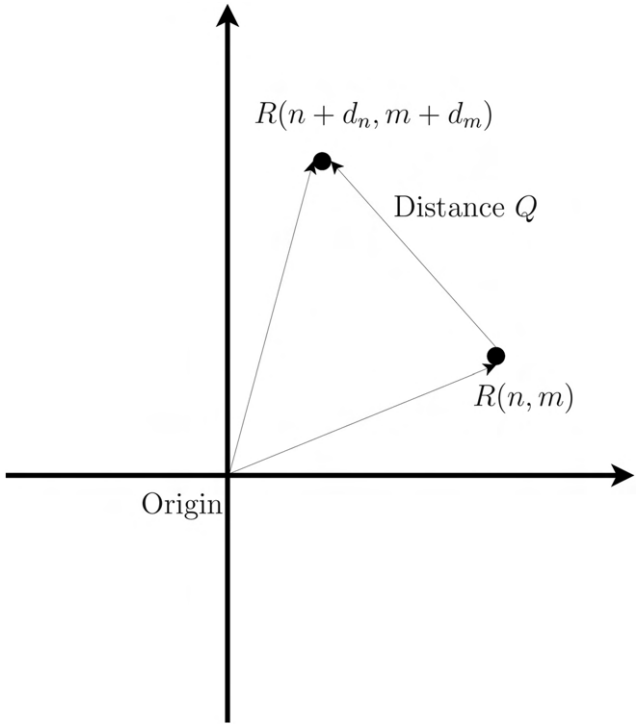
(*Appendices follow*)

Appendix A

Illustrations of Spatial Dependence and Semivariogram

Figure A1

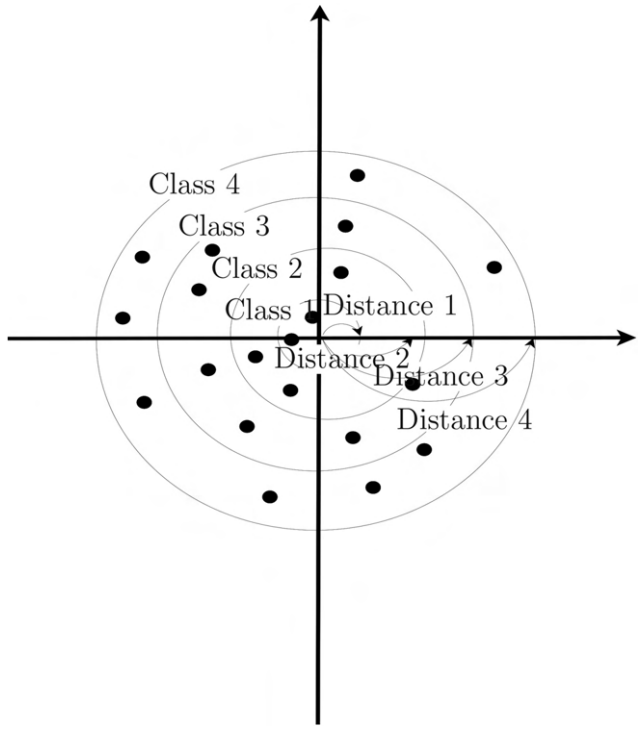
Distance Q Between the Two Data Points of Residuals R on x, y Coordinates



Note. Black dots indicate data points on *x*, *y* coordinates.

Figure A2

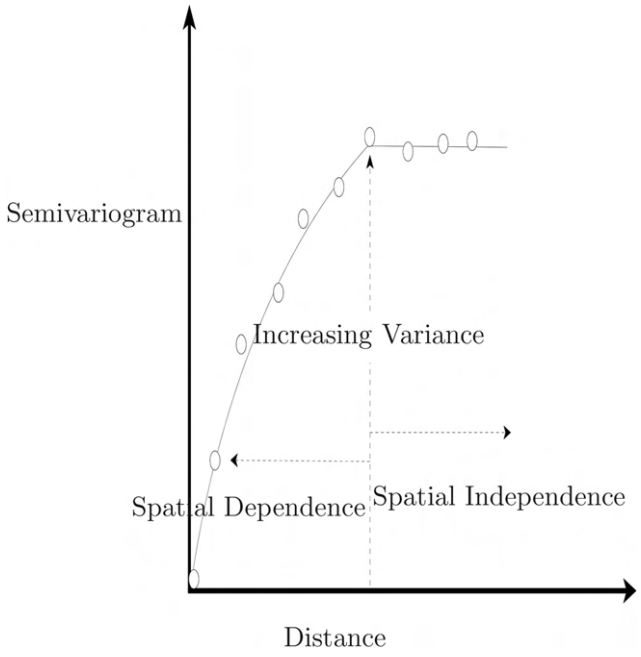
Classes for the Pairs of Data Points on x, y Coordinates



Note. Black dots indicate data points on *x*, *y* coordinates.

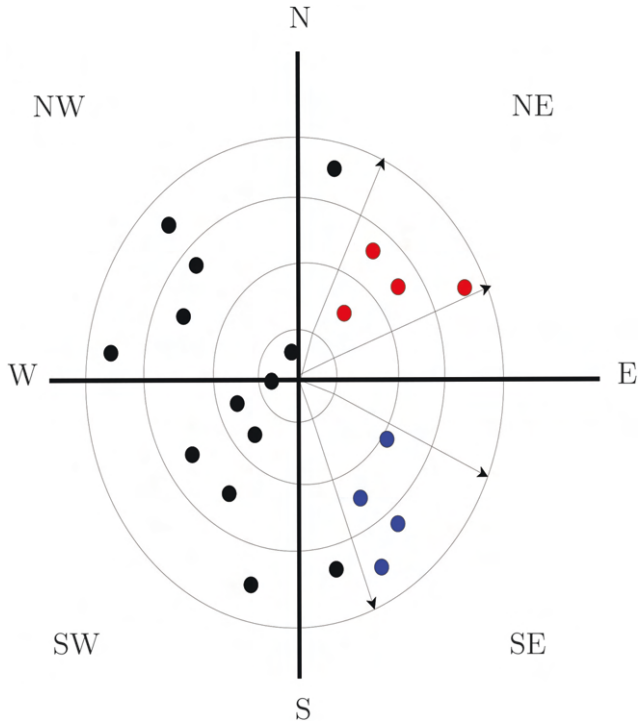
(*Appendices continue*)

Figure A3
Semivariogram Against Distance



Note. White dots indicate calculated semivariogram at each class.

Figure A4
Illustrations of the Isotrophy Assumption



Note. Dots indicate data points on *x*, *y* coordinates: Red dots indicate data points in the direction of northeast; Blue dots indicate data points in the direction of southeast. See the online article for the color version of this figure.

(*Appendices continue*)

Calculating Semivariogram Using R

An illustrative data set includes an observation identification index (Obs. ID), Pearson residuals, *x*-coordinate, and *y*-coordinate. The data set for 40 observations was generated with a standard normal distribution for the Pearson residuals and with selected values on *x*, *y* coordinates in pixels from the empirical study

Step 1 Calculate the Euclidean Distance and Create Classes

For 40 data points, there are 780 pairs of two data points. As an example, the Euclidean *distance* for a pair of ID = 1 on *x*, *y* coordinates of (1,268; 828) and ID = 2 on *x*, *y* coordinates of (1,304; 802) can be calculated as follows:

$$\sqrt{(1268 - 1304)^2 + (828 - 802)^2} = 44.407 \qquad (30)$$

Table A1
Illustrative Data

Obs. ID	Residuals	*x*-coordinate	*y*-coordinate	*R*(*n*, *m*) notation
1	−0.473	1,268	828	−0.473 (1,268,828)
2	−0.828	1,304	802	−0.828 (1,304,802)
3	1.585	1,317	807	1.585 (1,317,807)
4	−0.895	1,275	198	−0.895 (1,275,198)
5	0.723	1,261	216	0.723 (1,261,216)
6	−0.768	1,289	177	−0.768 (1,289,177)
7	−0.120	1,286	184	−0.120 (1,286,184)
8	1.496	753	725	1.496 (753,725)
9	0.255	681	280	0.255 (681,280)
10	0.241	682	282	0.241 (682,282)
11	1.076	1,222	790	1.076 (1,222,790)
12	1.349	1,209	788	1.349 (1,209,788)
13	0.237	760	264	0.237 (760,264)
14	−0.669	773	276	−0.669 (773,276)
15	−0.989	648	182	−0.989 (648,182)
16	0.176	647	177	0.176 (647,177)
17	−0.345	1,262	249	−0.345 (1,262,249)
18	−0.920	821	193	−0.920 (821,193)
19	1.821	828	205	1.821 (828,205)
20	−0.268	1,224	700	−0.268 (1,224,700)
21	−1.147	1,231	712	−1.147 (1,231,712)
22	−0.072	785	703	−0.072 (785,703)
23	0.419	790	704	0.419 (790,704)
24	−2.914	940	616	−2.914 (940,616)
25	−0.364	1,270	259	−0.364 (1,270,259)
26	0.062	976	289	0.062 (976,289)
27	−0.737	698	210	−0.737 (698,210)
28	−0.574	701	212	−0.574 (701,212)
29	−0.357	1,321	166	−0.357 (1,321,166)
30	0.324	1,319	180	0.324 (1,319,180)
31	0.028	1,264	891	0.028 (1,264,891)
32	−1.038	1,263	881	−1.038 (1,263,881)
33	0.494	1,254	830	0.494 (1,254,830)
34	0.417	1,252	823	0.417 (1,252,823)
35	0.804	811	698	0.804 (811,698)
36	0.606	760	722	0.606 (760,722)
37	1.196	750	730	1.196 (750,730)
38	0.988	690	226	0.988 (690,226)
39	−0.538	693	210	−0.538 (693,210)
40	−1.273	1,159	176	−1.273 (1,159,176)

In R, the Euclidean distance for all pairs can be calculated using the `dist` function in the `stats` package:

```
data <- read.table("C:/Variogram_data.txt",
  header=T)
#Columns of 3 and 4 are x, y coordinates in the
  #imported data.
dists <- dist(data[,3:4],method="euclidean")
summary(dists) #summary of distance
```

The next step is to split distance values into classes to have enough data points to calculate the variance in Step 2. Based on the results of `summary(dists)`, 11 classes with min = 0 and max = 1,000 were created using the `breaks` function:

```
> summary(dists)
```

```
Min.   1st Qu.   Median   Mean     3rd Qu.   Max.
2.236  317.601   527.665  482.093  658.055   943.655
```

```
> breaks = seq(0, 1000; l = 11)
> breaks
 [1]  0 100 200 300 400 500 600 700 800 900 1000
```

Step 2 Calculate the Variance Using Equation 2

Based on the 11 classes from Step 1, distances Q^c (from the center of the bins; e.g., 50 for the interval [0, 100], 150 for the interval [101, 200], etc.) and N^c (the number of pairs within class *c*) in Equation 2 (of the main paper) can be calculated. The R function `variog` in the `geoR` package can be used to obtain Q^c, N^c, and the semivariogram for each class:

```
#variog function: computing omnidirectional
  #variogram
library(geoR)
v <- variog(coords = data[,3:4], data = data
  [,2],breaks = breaks)
v$u #Distance Q
v$n #Num. of pairs: N_c
v$v #estimated variance
```

Table A2
Results of Illustrative Data

Class ID	Distance Q_c	Num. of pairs N_c	Estimated variance
1	50	119	0.6084346
2	150	58	1.0680901
3	250	16	1.9122761
4	350	19	2.7045106
5	450	121	1.1328401
6	550	160	0.8262266
7	650	129	0.7123553
8	750	96	0.7252405
9	850	54	0.7128542
10	950	8	0.7610677

(*Appendices continue*)

Table A3
Results of the Selected Pairs

Pair ID	Residual for a Pair	Variance
15,2	–0.989, −0.828	$(-0.989 + 0.828)^2 = 0.026$
16,2	0.176, −0.828	$(0.176 + 0.828)^2 = 1.008$
15,3	–0.989, 1.585	$(-0.989 - 1.585)^2 = 6.625$
16,3	0.176, 1.585	$(0.176 - 1.585)^2 = 1.985$
31,15	0.028, −0.989	$(0.028 + 0.989)^2 = 1.034$
32,15	–1.038, −0.989	$(-1.038 + 0.989)^2 = 0.002$
31,16	0.028, 0.176	$(0.028 - 0.176)^2 = 0.022$
32,16	–1.038, 0.176	$(-1.038 - 0.176)^2 = 1.474$
		sum = 1.522

Results of the function `variog` are summarized as follows:

To illustrate the calculation of semivariogram estimate = .761 using Equation 2, we chose Class = 10 (distance interval [901,1,000]) having $N_c = 8$ using the following code:

```
dists <- dist(data[,3:4],method="euclidean")
summary(dists)
dists <- as.matrix(dists,nrow = 40,ncol = 40)
which(dists > 901, arr.ind = T)
```

The selected pairs and their residuals are as follows:

The variance for Class = 10 can be calculated using Equation 2 as follows:

$$\begin{aligned}\hat{\gamma}(Q_c = 950) = \frac{1}{8}\{&(-0.989 + 0.828)^2 + (0.176 + 0.828)^2 \\ &+ (-0.989 - 1.585)^2 + (0.176 - 1.585)^2 + (0.028 + 0.989)^2 \\ &+ (-1.038 + 0.989)^2 + (0.028 - 0.176)^2 + (-1.038 - 0.176)^2\} \\ &= 1.522\end{aligned} \tag{31}$$

The semivariogram is $0.5 \times \hat{\gamma}(Q_c = 950) = 0.5 \times 1.522 = 0.761$. The same process can be replicated for the other classes.

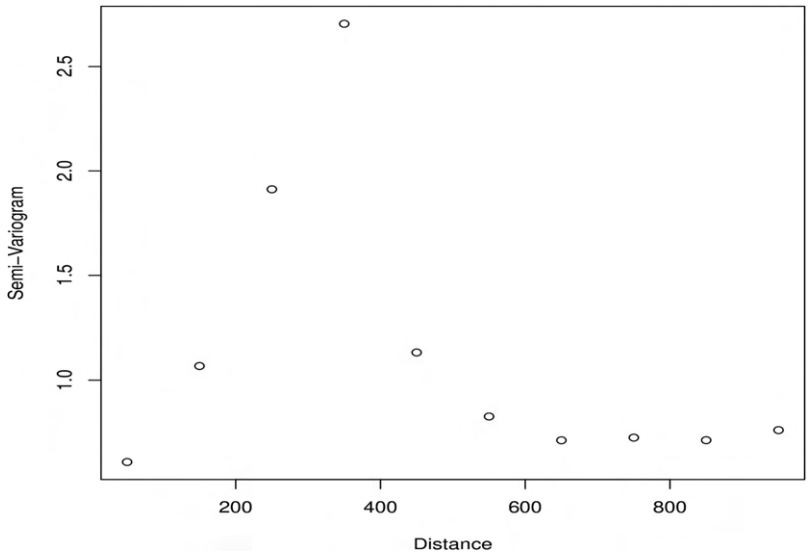
Step 3

Plot the estimated variance for all classes against distance. A figure of semivariogram can be created in R:

```
plot(v$u, v$v, xlab="Distance", ylab="Semi-
  Variogram")
```

Figure A5 shows that the value of semivariogram peaks at the distance of 350 and then declines, indicating that, in this example, data points that are spread further apart become more similar.

Figure A5
Plot of Semivariogram



(*Appendices continue*)

Appendix B

Implementation in the mgcv Package

Below, we describe how random effects and smooth functions in Equation 17 can be specified in the mgcv package.

As discussed earlier, the random effects are equivalent to a smooth with penalty matrix *I*. The random intercept parameters in Equation 17 (θ_j, ζ_i) can be specified as follows:

```
bam(y ~ s(x, bs="re")).
```

In the above expression, the function s indicates the use of a smooth function for the covariate x (i.e., an identification variable for persons for θ_j and an identification variable for items for ζ_i) and the argument bs="re" indicates the selected smoothing basis re.

Nonlinear relation between an outcome variable y and a covariate x is specified as follows:

```
bam(y ~ s(x, bs="cr", k=)).
```

In the expression, the argument bs = "cr" indicates the selected smoothing basis CRS, and the argument k = indicates the number of basis functions.

A smooth version of random slope in GLMM can be specified by a random smooth to model by-group variation in nonlinear effects. For a binary time-varying AR(1) effect (that is, the effect of $y_{(t-1)lji}$), the ordered factor of $y_{(t-1)lji}$ was considered and the ordered factor includes the fixed intercept of the AR(1) effect and reference smooth:

```
data$OFylag1 <- as.factor(data$ylag1)
data$OFylag1 <- as.ordered(data$OFylag1)
contrasts(data$OFylag1) <- 'contr.treatment'
bam(y ~ OFylag1 + s(time, bs="cr", k=) + s(time, by=OFylag1, bs="cr", k=)).
```

Because the smooth with the ordered factor is centered, the fixed intercept term OFylag1 is needed to avoid artifacts due to the centering constraint (van Rij et al., 2019). The fixed intercept effect of OFylag1 is the mean difference between consecutive time points with same values ((0, 0) or (1, 1)) and those with opposite values ((0, 1) or (1, 0)) (on the logit scale).

For numeric spatial lag covariates ($d_{(t-1)lji}$ and $c_{(t-1)lji}$) that vary across time, the spatial lag covariates are group and time is x in the mgcv package:

```
bam(y ~ s(x, by=group, bs="cr", k=)).
```

Here, the by=group argument is used to multiply the smooth function for x by covariates group.

For the two-dimensional nonlinear interaction between trial and time in different units, the tensor product smooth interaction can be included as ti(trial,time) in the mgcv package. Note that ti is used instead of te to model a pure interaction term, as in the ANOVA:

```
bam(y ~ ti(trial, time, bs="cr", k=)).
```

For the two-dimensional nonlinear interaction between xposition and yposition in the same units, the TPRS can be set as:

```
bam(y ~ s(xposition, yposition, bs="tp", k=)).
```

When the TPRS is set as an alternative to the CRS, the argument bs = "cr" can be replaced with the argument bs = "tp."

(*Appendices continue*)

Appendix C

R Code for a Generalized Additive Logistic Regression Model

The syntax file below is used to fit a generalized additive logistic regression model.

```
library(gstat) #for bubble functions
library(sp)    #for a coordinates function
library(mgcv)  #for a bam function
library(itsadug) #for bam result interpretations
library(Hmisc) #for Somers' rank correlation
library(ggplot2) #for Figure 3 plots

############################ Temporal (Auto)Correlations for trials by sessions

#The empirical data set (called "sleep2_condition3") can be found in Open Science Framework website.
#It is added under a C directory in this example.
data <- read.table("C:/sleep2_condition3.txt",header=T)
attach(data)
data <- data[order(time,trialseq),] # Sorts the rows of the data, first by time, second by trialseq.
head(data)
#the unique values for each variable:
unique_time <- sort(unique(time))
unique_trialseq <- sort(unique(trialseq))
unique_person <- sort(unique(person))
unique_itemid <- sort(unique(itemid))
unique_session <- sort(unique(session))
number_time <- length(unique_time)         # 132
number_trialseq <- length(unique_trialseq) # 64
number_person <- length(unique_person)     # 60
number_itemid <- length(unique_itemid)     # 8
number_session <- length(unique_session)   # 2
#initializing matrix and index used in following loop:
empirical_logit <- matrix(nrow=(number_time*number_trialseq*number_session),ncol=5)
index <- 1
for (t in 1:number_time){
for (l in 1:number_trialseq){
for (s in 1:number_session){
# For every unique combination of time, trialseq, and session:
data_tls <- data[(data$session==unique_session[s] & data$time==unique_time[t] & data$trialseq==unique_trialseq[l]),]
#the data for that time, trialseq, and session.
y_tls <- data_tls$y #the y values for that time, trialseq, and session (used to calculate empirical logit).
proportion <- sum(y_tls)/length(y_tls)   #the proportion calculated. Note that the length of y_tls is equal to number_person.
emp_logit <- log(proportion/(1 - proportion))   #the empirical logit calculated
empirical_logit[index,] <- c(unique_time[t], unique_trialseq[l], unique_session[s], proportion, emp_logit)
#adding unique combination of variables used and calculated values to matrix.
index <- index + 1 #the row index of the empirical_logit matrix to use, increases after every entry.
}}}

empirical_logit <- data.frame(empirical_logit) #changes the matrix into a data frame.
names(empirical_logit) <- c("time.el", "trialseq.el", "session.el", "proportion.el", "empirical_logit.el")
attach(empirical_logit)
head(empirical_logit)
empirical_logit <- empirical_logit[order(time.el),]
#changes the matrix created in the previous loop to a data frame, ordered by time for the next step.

time_lag_max <- 20 #change this value to change the maximum time lag which autocorrelations/partial autocorrelations use
autocorrelations <- array(dim=c(number_trialseq, time_lag_max, number_session))
partial_autocorrelations <- array(dim=c(number_trialseq, time_lag_max, number_session))

for (s in 1:number_session){
for (time_lag in 1:time_lag_max){
```

(*Appendices continue*)

```
for (l in 1:number_trialseq){
# For every unique combination of session, trialseq, and time lag (ranging from 1 to 20):
empirical_logit_ls <- empirical_logit.el[(unique_trialseq[l]==trialseq.el & unique_session[s]==session.el)]
#obtains the empirical logit values for the trialseq and session (note they're already ordered by time).
ac <- acf(empirical_logit_ls, lag.max = time_lag, type=c("correlation"), plot=FALSE) #calculates the autcorrelation
autocorrelations[l,time_lag,s] <- as.numeric(unlist(ac)[time_lag + 1]) #adds the autocorrelation to the array.
pac <- pacf(empirical_logit_ls, lag.max = time_lag, type=c("correlation"), plot=FALSE) #calculates the partial autocorrelation
partial_autocorrelations[l,time_lag,s] <- as.numeric(unlist(pac)[time_lag]) #adds the partial autocorrelation to the array.
}}}

#creates a dataframe containing results for each session:
#autocorrelations for session 1 and session 2:
(autocorrelations_session_1 <- data.frame(autocorrelations[,,1]))
(autocorrelations_session_2 <- data.frame(autocorrelations[,,2]))
# Partial autocorrelations for session 1 and session 2:
(partial_autocorrelations_session_1 <- data.frame(partial_autocorrelations[,,1]))
(partial_autocorrelations_session_2 <- data.frame(partial_autocorrelations[,,2]))

##### R code for generating Figure 3:
#empirical logit (trend) plots:
time_breaks <- seq(0,1500,by=100) # Tick marks for x-axis (ranging from minimum to maximum time in equal intervals).

# Figure 3 (top left):
els1 <- empirical_logit[(empirical_logit$session.el==1),] #empirical logits for session 1
els1$time.el <- as.factor(els1$time.el)
ggplot(els1, aes(x=time.el, y=empirical_logit.el)) +
geom_boxplot() +
labs(title = "Session 1", x = "Time (ms)", y = "Empirical Logit for Trials") +
scale_x_discrete(breaks=time_breaks) +
theme_bw()

# Figure 3 (top right)
els2 <- empirical_logit[(empirical_logit$session.el==2),] #empirical logits for session 2
els2$time.el <- as.factor(els2$time.el)
ggplot(els2, aes(x=time.el, y=empirical_logit.el)) +
geom_boxplot() +
labs(title = "Session 2", x = "Time (ms)", y = "Empirical Logit for Trials") +
scale_x_discrete(breaks=time_breaks) +
theme_bw()

#data frames for autocorrelation and partial autocorrelation plots:
ac1 <- ac2 <- pac1 <- pac2 <- matrix(nrow=number_trialseq*time_lag_max,ncol=3)
count <- 1
for (l in 1:number_trialseq){
for (t in 1:time_lag_max){
ac1[count,] <- c(unique_trialseq[l],t,autocorrelations_session_1[l,t]) # Autocorrelations for session 1
ac2[count,] <- c(unique_trialseq[l],t,autocorrelations_session_2[l,t]) # Autocorrelations for session 2
pac1[count,] <- c(unique_trialseq[l],t,partial_autocorrelations_session_1[l,t]) # Partial autocorrelations for session 1
pac2[count,] <- c(unique_trialseq[l],t,partial_autocorrelations_session_2[l,t]) # Partial autocorrelations for session 2
count <- count + 1
}}
ac1 <- data.frame(ac1)
ac2 <- data.frame(ac2)
pac1 <- data.frame(pac1)
pac2 <- data.frame(pac2)
names(ac1) <- names(ac2) <- names(pac1) <- names(pac2) <- c("trialseq.ac","time_lag.ac","correlation.ac")

#autocorrelation and partial autocorrelation plots:
# Figure 3 (middle left):
ac1$time_lag.ac <- as.factor(ac1$time_lag.ac)
ggplot(ac1, aes(x=time_lag.ac, y=correlation.ac)) +
geom_boxplot() +
labs(title = "Session 1", x = "Time Lag", y = "Autocorrelation for Trials") +
theme_bw()
```

(*Appendices continue*)

```
# Figure 3 (middle right):
ac2$time_lag.ac <- as.factor(ac2$time_lag.ac)
ggplot(ac2, aes(x=time_lag.ac, y=correlation.ac)) +
geom_boxplot() +
labs(title = "Session 2", x = "Time Lag", y = "Autocorrelation for Trials") +
theme_bw()
# Figure 3 (bottom left):
pac1$time_lag.ac <- as.factor(pac1$time_lag.ac)
ggplot(pac1, aes(x=time_lag.ac, y=correlation.ac)) +
geom_boxplot() +
labs(title = "Session 1", x = "Time Lag", y = "Partial Autocorrelation for Trials") +
theme_bw()
# Figure 3 (bottom right):
pac2$time_lag.ac <- as.factor(pac2$time_lag.ac)
ggplot(pac2, aes(x=time_lag.ac, y=correlation.ac)) +
geom_boxplot() +
labs(title = "Session 2", x = "Time Lag", y = "Partial Autocorrelation for Trials") +
theme_bw()


############################## Spatial Correlations
#creat "person" and "itemid" as factors in R
data$person <- as.factor(data$person)
data$itemid <- as.factor(data$itemid)

#fit a null model
null <- bam(y ~ 1 + s(person,bs="re") + s(itemid,bs="re"), data=data, family=binomial, method="ML")

#calculate Pearson residuals for the null model
residuals <- residuals(null,type="pearson")

#extract x,y coordinates from the data
data$x <- data$xposition
data$y <- data$yposition

#create a data set having Pearson residuals and x,y coordinates
mydata <- data.frame(residuals,data$x,data$y)

#assine x,y coordinates in the dataset to "coordinates" variables
coordinates(mydata) <- c("data.x","data.y")

#create Figure 4 (top)
vari <- variogram(residuals ~ 1, data=mydata)
plot(vari)

#check the isotropy assumption
vari <- variogram(residuals ~ 1, data=mydata, alpha=c(0,45,90,135))

#create Figure 5
bubble(mydata,"residuals",col=c("black","grey"),main = "Pearson Residuals",
xlab="x-coordinates (ranging from -1517 to 2943)", ylab="y-coordinates (ranging from -818 to 2136)")


############################## Spatial by Temporal Interactions

#sleep2_condition3_na.omit1.txt does not have lag effects
#importing data
data <- read.table("C:/sleep2_condition3_na.omit1.txt",header=T)

#create variables as factors in R
data$person <- as.factor(data$person)
data$itemid <- as.factor(data$itemid)
data$session <- as.factor(data$session)
data$timeblock <- as.factor(data$timeblock)
```

(*Appendices continue*)

```
#fit a null model by timeblock
null.1 <- bam(y ~ 1 + s(person,by=timeblock,bs="re") + s(itemid,by=timeblock,bs="re"), data=data, family=binomial, method="ML")

#calculate Pearson residuals for the null model
residuals <- residuals(null.1,type="pearson")

#extract x,y coordinates from the data
data$x <- data$xposition
data$y <- data$yposition

#create a data set having Pearson residuals, x,y coordinates, and timeblock variables
mydata <- data.frame(residuals,data$x,data$y,data$timeblock)

#assine x,y coordinates in the dataset to "coordinates" variables
coordinates(mydata) <- c("data.x","data.y")


#create variogram by timeblocks
timeblock6 <- subset(mydata, data.timeblock=="6") #A timeblock variable goes from 1 to 6

#create Figure 4 (bottom)
vari <- variogram(residuals ~ 1, data=timeblock6)
plot(vari)



############################ Model Selection

#create variables as factors in R
data$person <- as.factor(data$person)
data$itemid <- as.factor(data$itemid)
data$session <- as.factor(data$session)

#create a "OFylag1" variable as ordered factors in R
data$OFylag1 <- as.factor(data$ylag1)
data$OFylag1 <- as.ordered(data$OFylag1)
contrasts(data$OFylag1) <- 'contr.treatment'

#fit candidate models regarding spatial and temporal effects
null <- bam(y ~ 1 + s(person,bs="re") + s(itemid,bs="re"), data=data, family=binomial, method="ML")

modelA <- bam(y ~ 1 + OFylag1 + s(time,by=OFylag1,bs="cr",k=10) + s(time,bs="cr",k=10) +
s(person,bs="re") + s(itemid,bs="re"), data=data, family=binomial, method="ML")

modelB <- bam(y ~ 1 + OFylag1 + s(time,by=OFylag1,bs="cr",k=10) + s(time,bs="cr",k=10) +
s(trialseq, bs="cr",k=10) +  ti(trialseq,time,bs="cr",k=10) +  s(person,bs="re") + s(itemid,bs="re"),
data=data, family=binomial, method="ML")

modelC <- bam(y ~ 1 + OFylag1 + s(time,by=OFylag1,bs="cr",k=10) + s(time,by=t_dlag1,bs="cr",k=10) +
s(time,by=c_dlag1,bs="cr",k=10) + s(time,bs="cr",k=10) + s(trialseq, bs="cr",k=10) +
ti(trialseq,time,bs="cr",k=10) +  s(person,bs="re") + s(itemid,bs="re"),
data=data, family=binomial, method="ML")

#modelD has a convergence problem.
#modelD <- bam(y ~ 1 + OFylag1 + s(time,by=OFylag1,bs="cr",k=10) + s(time,by=t_dlag1,bs="cr",k=10) +
s(time,by=c_dlag1,bs="cr",k=10) + s(time,bs="cr",k=10) + s(trialseq, bs="cr",k=10) +
ti(trialseq,time,bs="cr",k=10) +  s(xposition,yposition,bs="tp",k=10) + s(person,bs="re") + s(itemid,bs="re"), data=data, family=bi


#By default this routine uses a definition of the effective degrees of freedom that includes smoothing parameter uncertainty
c(logLik.gam(null),logLik.gam(modelA),logLik.gam(modelB),logLik.gam(modelC))
c(BIC(null),BIC(modelA),BIC(modelB),BIC(modelC))
c(AIC(null),AIC(modelA),AIC(modelB),AIC(modelC))
```

(*Appendices continue*)

```
############################ Model Fit and Evaluation

#modelE = modelC + experimental conditions
modelE <- bam(y ~ talker_e*session_e*sleep1 + talker_e*session_e*sleep2 + OFylag1 +
s(time,by=OFylag1,bs="cr",k=10) + s(time,by=t_dlag1,bs="cr",k=10) + s(time,by=c_dlag1,bs="cr",k=10) +
s(time,bs="cr",k=10) + s(trialseq,bs="cr",k=10) + ti(trialseq,time,bs="cr",k=10) +  s(person,bs="re") +
s(itemid,bs="re"), data=data, family=binomial, method="ML")
summary(modelE)


#check computational time
system.time(modelE <- bam(y ~ talker_e*session_e*sleep1 + talker_e*session_e*sleep2 + OFylag1 +
s(time,by=OFylag1,bs="cr",k=10) + s(time,by=t_dlag1,bs="cr",k=10) + s(time,by=c_dlag1,bs="cr",k=10) +
s(time,bs="cr",k=10) + s(trialseq,bs="cr",k=10) + ti(trialseq,time,bs="cr",k=10) +  s(person,bs="re") +
s(itemid,bs="re"), data=data, family=binomial, method="ML"))

#extract estimates of modelE
gamma <- modelE$coefficients
gamma <- matrix(gamma)

#extract the covariance matrix of the estimates
var.cov <- vcov(modelE,freq=T)
var.cov


#convert smoothing parameters to variance of random effects
#The first column of the matrix gives standard deviations for each
#term, while the subsequent columns give lower and upper confidence
#bounds, on the same scale.
var.random <- gam.vcomp(modelE)

#plot smooth functions of modelE
plot(modelE)
plot(modelE, shade = TRUE, pages = 1, scale = 0)


#time-varying AR
plot(modelE, select = 1)

#target distance AR
plot(modelE, select = 2)

#competitor distance AR
plot(modelE, select = 3)

#time
plot(modelE, select = 4)

#trial
plot(modelE, select = 5)

#interaction between time and trial
plot(modelE, select = 6)

#person random effect
plot(modelE, select = 7)

#item random effect
plot(modelE, select = 8)
```

(*Appendices continue*)

```
#Model evaluation
#extract the fitted value of modelE
gam.check(modelE,type="pearson")
fit <- fitted(modelE)

#calculate Pearson residuals of modelE and plot the fitted values against residuals
res <- residuals(modelE,type="pearson")
hist(res, xlab="Pearson Residuals")
plot(fit,res)

#calculating Somers' rank correlation
somers2(fit, as.numeric(na.omit(data$y)))


#fit a model which ignore temporal and spatial effects
modelF <- bam(y ~ talker_e*session_e*sleep1 + talker_e*session_e*sleep2 +  s(person,bs="re") +
s(itemid,bs="re"), data=data, family=binomial, method="ML")
summary(modelF)




############################## Other Modeling Considerations

#Spatial smoothing (described under a summary and discussion section

spatial <- bam(y ~ s(xposition,yposition,bs="tp",k=100), data=data, family=binomial, method="ML")
pvisgam(spatial)

vis.gam(spatial,view=c("xposition","yposition"),plot.type="contour",color="terrain")


#Experimental condition effects

data$OFylag1 <- as.factor(data$ylag1)
data$OFylag1 <- as.ordered(data$OFylag1)
contrasts(data$OFylag1) <- 'contr.treatment'

data$session <- data$session - 1
data$OFsession <- as.factor(data$session)
data$OFsession <- as.ordered(data$OFsession)
contrasts(data$OFsession) <- 'contr.treatment'


modelG <- bam(y ~ 1 + OFylag1 + OFsession  + s(time,by=OFylag1,bs="cr",k=10) + s(time,by=t_dlag1,bs="cr",k=10) +
s(time,by=c_dlag1,bs="cr",k=10) + s(time,bs="cr",k=10) + s(trialseq,by=OFsession,bs="cr",k=10) +
s(trialseq, bs="cr",k=10) +  ti(trialseq,time,bs="cr",k=10) +  s(person,bs="re") + s(itemid,bs="re"),
data=data, family=binomial, method="ML")

modelH <- bam(y ~ 1 + OFylag1 + s(time,by=OFylag1,bs="cr",k=10) + s(time,by=t_dlag1,bs="cr",k=10) +
s(time,by=c_dlag1,bs="cr",k=10) + s(time,bs="cr",k=10) + s(trialseq, bs="cr",k=10) +
ti(trialseq,time,bs="cr",k=10) +  s(time,person,bs="fs",m=1) + s(time,itemid,bs="fs",m=1),
data=data, family=binomial, method="ML")

plot(modelH, select=8, rug=F)
plot(modelH, select=9, rug=F)
```

(*Appendices continue*)

Appendix D

Processing and Coding of the Eye-Tracking Data

The Eyelink 1000 (SR Research) eye-tracker used in this substantive example outputs the data in a format that indicates at each millisecond whether the eyes were fixating, saccading, or blinking. Fixations are indexed by the *x*, *y* screen coordinates of the fixation point. Saccades are indicated by the *x*, *y*, coordinates of the starting and ending point. Based on prior research in this substantive area, we recoded the time associated with each saccade as a fixation to the object that the eye landed on at the end of the saccade, based on the idea that the cognitive processing of the to-be-fixated object began as or before the saccade was initiated (see Irwin, 2004; McMurray et al., 2009; McMurray et al., 2010). The time associated with blinks are similarly attributed to the object that the person fixated on after the blink. We then downsampled the data into units of 10 ms for computational efficiency, and because for the substantive research question we are addressing, 10-ms resolution is sufficient. The eye movement data record is synched with the audio recordings played to the participants using Matlab and the Psychophysics toolbox (Brainard & Vision, 1997). For each of the trials included in the analysis, we then extracted the fixation data from 180 ms–1,499 ms following the onset of the critical word, for example, "back." This results in a data matrix that indicates, at each 10-ms time point between 180 ms–1,499 ms, the *x*, *y* coordinates of the fixation point relative to the *x*, *y* coordinates of the computer screen. From this matrix, we then generated a binary code for each point in time indicating whether or not the participant was fixating on the target. We also generated, for each point in time, the distance from the current fixation point to the centroid of the target, and separately the distance to the centroid of the competitor.

References

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Irwin, D. E. (2004). Fixation location and fixation duration as indices of cognitive processing. *The Interface of Language, Vision, and Action: Eye Movements and the Visual World*, *217*, 105–133.

McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, *60*(1), 1–39.

Received October 22, 2019
Revision received May 19, 2021
Accepted September 7, 2021 ■