

BAYESIAN SEMIPARAMETRIC LONGITUDINAL INVERSE-PROBIT MIXED MODELS FOR CATEGORY LEARNING

MINERVA MUKHOPADHYAY

INDIAN INSTITUTE OF TECHNOLOGY

JACIE R. MCHANEY AND BHARATH CHANDRASEKARAN

NORTHWESTERN UNIVERSITY

ABHRA SARKAR 

UNIVERSITY OF TEXAS AT AUSTIN

Understanding how the adult human brain learns novel categories is an important problem in neuroscience. Drift-diffusion models are popular in such contexts for their ability to mimic the underlying neural mechanisms. One such model for gradual longitudinal learning was recently developed in Paulon et al. (J Am Stat Assoc 116:1114–1127, 2021). In practice, category response accuracies are often the only reliable measure recorded by behavioral scientists to describe human learning. Category response accuracies are, however, often the only reliable measure recorded by behavioral scientists to describe human learning. To our knowledge, however, drift-diffusion models for such scenarios have never been considered in the literature before. To address this gap, in this article, we build carefully on Paulon et al. (J Am Stat Assoc 116:1114–1127, 2021), but now with latent response times integrated out, to derive a novel biologically interpretable class of ‘inverse-probit’ categorical probability models for observed categories alone. However, this new marginal model presents significant identifiability and inferential challenges not encountered originally for the joint model in Paulon et al. (J Am Stat Assoc 116:1114–1127, 2021). We address these new challenges using a novel projection-based approach with a symmetry-preserving identifiability constraint that allows us to work with conjugate priors in an unconstrained space. We adapt the model for group and individual-level inference in longitudinal settings. Building again on the model’s latent variable representation, we design an efficient Markov chain Monte Carlo algorithm for posterior computation. We evaluate the empirical performance of the method through simulation experiments. The practical efficacy of the method is illustrated in applications to longitudinal tone learning studies.

Key words: category learning, B-splines, drift-diffusion models, functional models, inverse Gaussian distributions, longitudinal mixed models, speech learning.

Scientific Background Categorization decisions are important in almost all aspects of our lives—whether it is a friend or a foe, edible or non-edible, the word /bat/ or /hat/, etc. The underlying cognitive dynamics are being actively studied through extensive ongoing research (Glimcher & Fehr, 2013; Gold & Shadlen, 2007; Heekeren et al., 2004; Purcell, 2013; Schall, 2001; Smith & Ratcliff, 2004).

In typical multi-category decision tasks, the brain accumulates sensory evidence in order to make a categorical decision. This accumulation process is reflected in the increasing firing rates at local neural populations associated with different decisions. A decision is taken when neural activity in one of these populations reaches a particular threshold level. The decision category that is finally chosen is the one whose decision threshold is crossed first (Brody & Hanks, 2016;

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11336-024-09947-8>.

Correspondence should be made to Abhra Sarkar, Department of Statistics and Data Sciences, University of Texas at Austin, 105 East 24th Street D9800, Austin, TX78712, USA. Email: abhra.sarkar@utexas.edu

Gold & Shadlen, 2007). Changes in evidence accumulation rates and decision thresholds can be induced by differences in task difficulty and/or cognitive function (Cavanagh et al., 2011; Ding & Gold, 2013). Decision-making is also regulated by demands on both the speed and accuracy of the task (Bogacz et al., 2010; Milosavljevic et al., 2010).

Understanding the brain activity patterns for different decision alternatives is a key scientific interest in modeling brain mechanisms underlying decision-making. Statistical approaches with biologically interpretable parameters that further allow probabilistic clustering of the parameters (Lau & Green, 2007; Wade, 2023) associated with different competing choices can facilitate such inference, the parameters clustering together indicating similar behavior and difficulty levels.

Drift-Diffusion Models A biologically interpretable joint model for decision response accuracies and associated response times is obtained by imitating the underlying evidence accumulation mechanisms using latent drift-diffusion processes racing toward their respective boundaries, the process reaching its boundary first producing the final observed decision and the time taken to reach this boundary giving the associated response time (Fig. 1a) (Usher & McClelland, 2001).

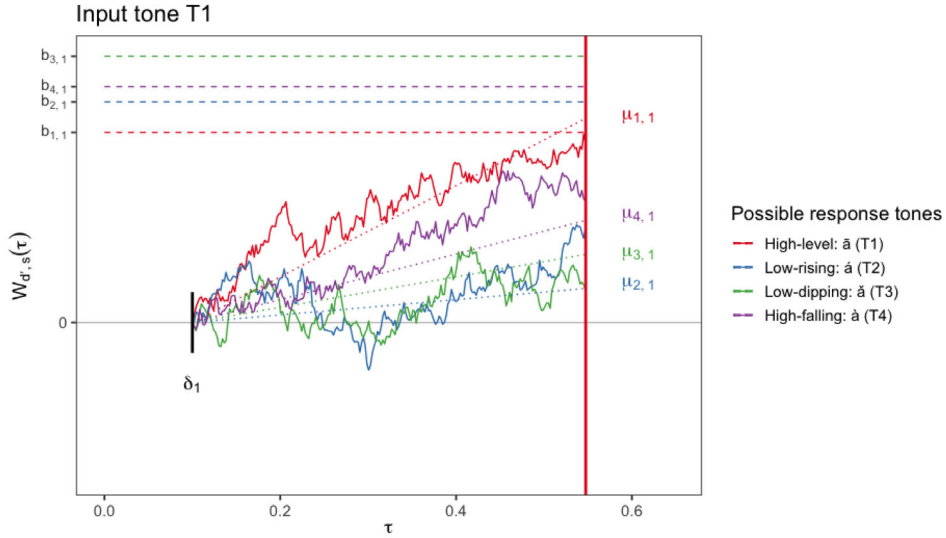
The literature on drift-diffusion processes for decision-making is rather vast but is mostly focused on simple binary decision scenarios with a single latent diffusion process with two boundaries, one for each of the two decision alternatives (Ratcliff, 1978; Ratcliff et al., 2016; Ratcliff & Rouder, 1998; Ratcliff & McKoon, 2008; Smith & Vickers, 1988). Multi-category drift-diffusion models with multiple latent processes are mathematically more easily tractable (Brown & Heathcote, 2008; Dufau et al., 2012; Kim et al., 2017; Leite & Ratcliff, 2010; Usher & McClelland, 2001) but the literature is sparse and focused only on simple static designs.

Learning to make categorization decisions is, however, a dynamic process, driven by perceptual adjustments in our brain and behavior over time. Category learning is thus often studied in longitudinal experiments. To address the need for sophisticated statistical methods for such settings, Paulon et al. (2021) developed an inverse Gaussian distribution-based multi-category longitudinal drift-diffusion mixed model.

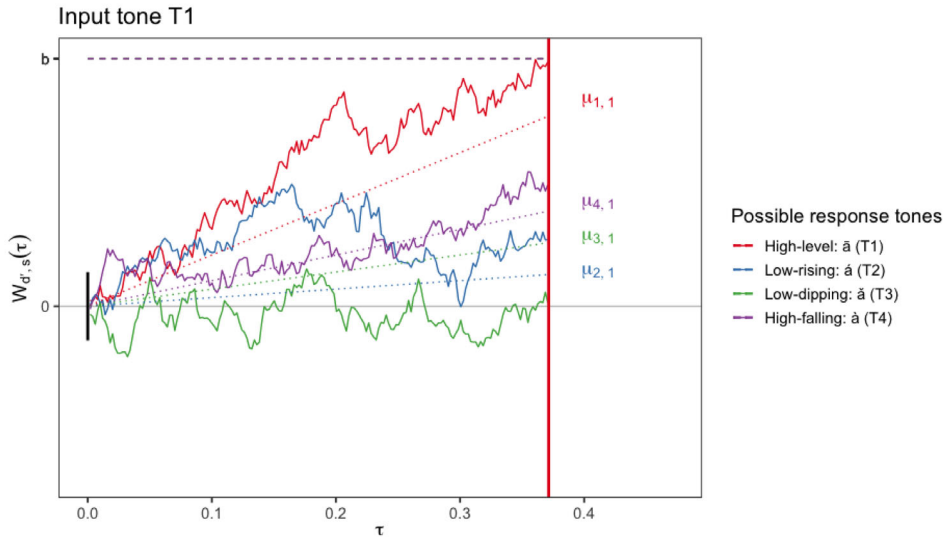
Data Requirements and Related Challenges Crucially, measurements on both the final decision categories and the associated response times are needed to estimate the drift and the boundary parameters from conventional drift-diffusion models, including the work by Paulon et al. (2021). Unfortunately, however, researchers often only record the participants' decision responses as their go-to measure of categorization performance, ignoring the response times (Chandrasekaran et al., 2014; Filoteo et al., 2010). Additionally, eliciting accurate response times can be methodologically challenging, e.g., in the case of experiments conducted online, especially during the Covid-19 pandemic (Roark et al., 2021), or when the response times from participants/patients are unreliable due to motor deficits (Ashby et al., 2003). Participants may also be asked to delay the reporting of their decisions so that delayed physiological responses that relate to decision-making can be accurately measured (McHaney et al., 2021). In such cases, the reported response times may not accurately relate to the actual decision times and hence cannot be used in the analysis. As a result, conventional drift-diffusion analysis that requires data on both response accuracies and response times, such as Paulon et al. (2021), cannot be used in such scenarios.

The Research Question The main research question addressed in this article is to see if a new class of drift-diffusion models can be designed for such scenarios which will allow the biologically interpretable drift-diffusion process parameters to be meaningfully recovered from data on input-output category combinations alone.

The Inverse-Probit Model Categorical probability models that build on latent drift-diffusion processes can be useful in providing biologically interpretable inference in data sets comprising input-output categories but no response times. To our knowledge, however, the problem has never been considered in the literature before. We aim to address this remarkable gap in this article.



(a) Drift-diffusion processes for tone learning when data on both response categories and response times are available.



(b) Drift-diffusion processes for tone learning proposed in this article – with additional identifiability restrictions imposed to accommodate scenarios when only data on response categories are available.

FIGURE 1.

Drift-diffusion model for tone learning. The tones $\{T1, T2, T3, T4\}$ represent the different categories; s denotes an input category, d' the different possible response categories, and d the final response category. Here we are illustrating a single trial with input tone T1 ($s = 1$) that was eventually correctly identified ($d = 1$). **a** Shows a process whose parameters can be inferred from data on both response categories and response times. Here, after an initial δ_s amount of time required to encode an input category s (here T1), the evidence in favor of different possible response categories d' accumulates according to latent Wiener diffusion processes $W_{d',s}(\tau)$ (red, blue, green, and purple) with drifts $\mu_{d',s}$. The decision d (here T1) is eventually taken if the underlying process (here the red one) reaches its decision boundary $b_{d,s}$. **b** shows a process with additional identifiability restrictions (for all d' and s , $\delta_s = 0$, $b_{d',s} = b$ fixed, and $\sum_{d'=1}^{d_0} \mu_{d',s} = d_0$) considered in this article which can be inferred from data on response categories alone.

By integrating out the latent response times from the joint inverse Gaussian drift-diffusion model for response categories and associated response times in Paulon et al. (2021), we can arrive at a natural albeit overparametrized model for the response categories. We refer to this as the ‘inverse-probit’ categorical probability model. This inverse-probit model serves as the starting point for the methodology presented in this article but, as we describe below, it also comes with significant and unique statistical challenges not encountered in the original drift-diffusion model. *Statistical Challenges* While scientifically desirable, unfortunately, it is also mathematically impossible to infer both the drifts and the boundaries in the inverse-probit model from data only on the decision accuracies. We must thus have to keep the values of either the drifts or the boundaries fixed and focus on inferring the other.

However, even when we fix either the drift or the decision boundaries, the problem of overparametrization persists. In the absence of response times, only the information on relative frequencies, that is empirical probabilities of taking a decision is available. As the total probability of observing any of the competing decisions is one, the identifiability problem remains for the chosen main parameters of interest, and appropriate remedial constraints need to be imposed.

Setting an arbitrarily chosen category as the reference provides a simple solution widely adopted in categorical probability models but comes with serious limitations, including breaking the symmetry of the problem, potentially making posterior inference sensitive to the specific choice of the reference category (Burgette & Nordheim, 2012; Johndrow et al., 2013).

By breaking the symmetry of the problem, a reference category also additionally makes it difficult to infer the potential clustering of the model parameters, especially across different panels. To see this, consider a problem with d_0 categories, with a logistic model for the probabilities $p_{s,d'} = \text{logistic}(\beta_{s,d'})$, $s, d' \in \{1 : d_0\}$, of choosing the d' th output category for the s th input category. For each input category s , by setting the s th output category as a reference, e.g., by fixing $\beta_{s,s} = 0$, one can then cluster the probabilities of incorrect decision choices, $p_{s,d'}, d' \neq s$. However, it is not clear how to compare the probabilities across different input categories (i.e., across the four panels in Fig. 2), e.g., how to test the equality of $p_{1,1}$ and $p_{2,2}$.

Finally, while coming up with solutions for the aforementioned issues, we must also take into consideration the complex longitudinal design of the experiments generating the data. Whatever strategy we devise, it should be amenable to a longitudinal mixed model analysis that ideally allows us to (a) estimate the smoothly varying longitudinal trajectories of the parameters as the participants learn over time, (b) accommodate participant heterogeneity, and (c) compare the estimates at different time points within and between different input categories.

Our Proposed Approach As a first step toward addressing the identifiability issues and related modeling challenges, we keep the boundaries fixed but leave the drift parameters unconstrained. The decision to focus on the drifts is informed by the existing literature on such models cited above where the drifts have almost always been allowed more flexibility. The analysis of Paulon et al. (2021) also showed that it is primarily the variations in the drift trajectories that explain learning while the boundaries remain relatively stable over time.

As a next step toward establishing identifiability, we apply a ‘sum to a constant’ condition on the drifts so that symmetry is maintained in the constrained model.

Implementation of this restriction brings in significant challenges. One possibility is to design a prior on the constraint space, a challenging task in itself. Additionally, posterior computation for such priors would also be extremely complicated in drift-diffusion models. Instead, we conduct inference with an unconstrained prior on the drift parameters and project the samples drawn from the corresponding posterior to the constrained space through a minimal distance mapping.

To adapt this categorical probability model to a longitudinal mixed model setting, we then assume that the drift parameters comprise input-response-category-specific fixed effects and subject-specific random effects, modeling them flexibly by mixtures of locally supported B-spline bases (de Boor, 1978; Eilers & Marx, 1996) spanning the length of the longitudinal experiment.

These effects are thus allowed to evolve flexibly as smooth functions of time (Morris, 2015; Ramsay & Silverman, 2007; Wang et al., 2016) as the participants get more experience and training in their assigned decision tasks.

We take a Bayesian route to estimation and inference. Carefully exploiting conditional prior-posterior conjugacy as well as our latent variable construction, we design an efficient Markov chain Monte Carlo (MCMC)-based algorithm for approximating the posterior, where sampling the latent response times for each observed response category greatly simplifies the computations.

We evaluate the numerical performance of the proposed approach in extensive simulation studies. We then apply our method to the PTC1 data set described below. These applications illustrate the utility of our method in providing insights into how the drift parameters characterize the rates of accumulation of evidence in the brain evolve over time, differ between input–output category combinations, as well as between individuals.

Differences from Previous Works This article differs in many fundamental ways from all existing works on drift-diffusion models, including Paulon et al. (2021), where response categories and response times were *both* observed and therefore the drift and boundary parameters could be modeled jointly with no identifiability issues. In contrast, the current work is motivated by scenarios where data on *only* response categories are available, leading us to the inverse-probit categorical probability model which, with its complex identifiability issues, brings in new unique challenges to performing statistical inference, confining us only to infer the drift parameters on a relative scale, achieved via a novel projection-based approach. The introduction and analysis of the inverse-probit model, addressing the significant new statistical challenges posed by it, ranging across (a) identifiability issues, (b) assessment of intra- and inter-panel similarities, (c) extension to complex longitudinal mixed effects settings to accommodate the motivating applications, (d) computational implementation of these new models, etc. are the novel contributions of this article.

Outline of the Article Section 1 describes our motivating tone learning study. Sections 2 and 3 develop our longitudinal inverse-probit mixed model. Section 4 outlines our computational strategies. Section 5 presents the results of simulation experiments. Section 6 presents the results of the proposed method applied to our motivating PTC1 study. Section 7 concludes the main article with a discussion. Additional details, including Markov chain Monte Carlo (MCMC)-based posterior inference algorithms, are deferred to the supplementary material.

1. The PTC1 Data Set

The PTC1 (pupillometry tone categorization experiment 1) data set is obtained from a Mandarin tone learning study conducted at the Department of Communication Science and Disorders, University of Pittsburgh (McHaney et al., 2021). Mandarin Chinese is a tonal language, which means that pitch patterns at the syllable level differentiate word meanings. There are four linguistically relevant pitch patterns in Mandarin that make up the four Mandarin tones: high-flat (Tone 1), low-rising (Tone 2), low-dipping (Tone 3), and high-falling (Tone 4). For example, the syllable /ma/ can be pronounced using the four different pitch patterns of the four tones, which would result in four different word meanings. Adult native English speakers typically experience difficulty differentiating between the four Mandarin tones because pitch contrasts at the syllable level are not linguistically relevant to word meanings in English (Wang et al., 1999, 2003). Thus, Mandarin tones are valid stimuli to examine how non-native speech sounds are acquired, which has implications for second language learning in adulthood. In PTC1, a group of native English-speaking younger adults learned to categorize monosyllabic Mandarin tones in a training task. During a single trial of training, an input tone was presented over headphones, and the participants were instructed to categorize the tone into one of the four tone categories via a button press on a keyboard. Corrective feedback in the form of “Correct” or “Wrong” was then provided

on screen. A total of $n = 28$ participants completed the training task across $T = 6$ blocks of training, each block comprising $L = 40$ trials. Figure 2 shows the middle 30% quantiles of the proportion of times the response to an input tone was classified into different tone categories over blocks across different subjects, each for the four input tones.

Pupillometry measurements were also taken during each trial. It is commonly used as a metric of cognitive effort during listening because increases in pupil diameter are associated with greater usage of cognitive resources (Parthasarathy et al., 2020; Peelle, 2018; Robison & Unsworth, 2019; Winn et al., 2018; Zekveld et al., 2011). One issue with pupillary responses however is that they unfold slowly over time. In view of that, unlike standard Mandarin tone training tasks, where the participants hear the input tone, press the keyboard response, and are provided feedback all within a few seconds (Chandrasekaran et al., 2016; Llanos et al., 2020; Reetzke et al., 2018; Smayda et al., 2015), in the PTC1 experiment, there was an intentional four-second delay from the start of the input tone to the response prompt screen where participants made their category decision via button press. This four-second delay allows the pupil to dilate in response to hearing the tone and begin to return to baseline before the participant makes a motor response to the button press. During this four-second period, participants have likely already made conscious category decisions. As such, the response times that are recorded in the end are not meaningful measures of their actual decision times.

This presents a critical limitation for using these response times for further analysis. Conventional drift-diffusion analysis that requires data on response times, such as the one presented in Paulon et al. (2021), can no longer be directly applied here. The focus of this article is to see if the drift-diffusion parameters can still be meaningfully recovered from input–output tone categories alone in the PTC1 data.

We found drift-diffusion analysis in the absence of reliable data on response times challenging enough to merit its separate treatment presented here. Relating drift-diffusion parameters to measures of cognitive effort such as pupillometry is another challenging problem that we are pursuing separately elsewhere.

2. Inverse-Probit Model

The starting point for the proposed inverse-probit categorical probability model follows straightforwardly by integrating out the (unobserved) response times from the joint model for response categories and associated response times developed in Paulon et al. (2021). The derivation of this original joint model illustrates its latent drift-diffusion process-based underpinnings (Fig. 1a). Later such construction will also be crucial in understanding the diffusion process-based foundations of the marginal categorical probability model modified with identifiability constraints proposed in this article (Fig. 1b). We therefore present the derivation from Paulon et al. (2021) ditto here which also keeps the main paper self-contained.

To begin with, a Wiener diffusion process $W(\tau)$ over domain $\tau \in (0, \infty)$ can be specified as $W(\tau) = \mu\tau + \sigma B(\tau)$, where $B(\tau)$ is the standard Brownian motion, μ is the drift rate, and σ is the diffusion coefficient (Cox & Miller, 1965; Ross et al., 1996). The process has independent normally distributed increments, i.e., $\Delta W(\tau) = \{W(\tau + \Delta\tau) - W(\tau)\} \sim \text{Normal}(\mu\Delta\tau, \sigma^2\Delta\tau)$, independently from $W(\tau)$. The first passage time of crossing a threshold b , $\tau = \inf\{\tau' : W(0) = 0, W(\tau') \geq b\}$, is then distributed according to an inverse Gaussian distribution (Chhikara, 1988; Lu, 1995; Whitmore & Seshadri, 1987) with mean b/μ and variance $b\sigma^2/\mu^3$.

Given a perceptual stimulus s and a set of decision choices $d' \in \{1 : d_0\}$, the neurons in the brain accumulate evidence in favor of the different alternatives. Modeling this behavior using latent Wiener processes $W_{d',s}(\tau)$ with unit variances, assuming that a decision d is made when the decision threshold $b_{d,s}$ for the d th option is crossed first, as illustrated in Fig. 1a, a probability

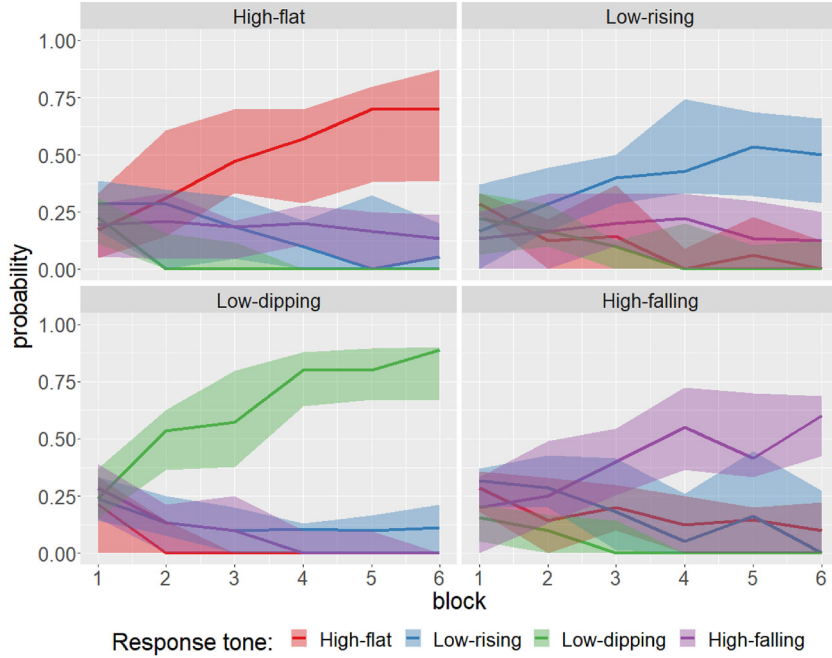


FIGURE 2.

Description of PTC1 data: The proportion of times the response to an input tone was classified into different tone categories over blocks across different subjects, each for the four input tones (indicated in the panel headers). The thick line represents the median performance and the shaded region indicates the corresponding middle 30% quantiles across subjects.

model for the time τ_d to reach decision d is obtained as

$$f(\tau_d | \delta_s, \mu_{d,s}, b_{d,s}) = \frac{b_{d,s}}{\sqrt{2\pi}} (\tau_d - \delta_s)^{-3/2} \exp \left[-\frac{\{b_{d,s} - \mu_{d,s}(\tau_d - \delta_s)\}^2}{2(\tau_d - \delta_s)} \right], \quad (1)$$

where $\mu_{d,s}$ denotes the rate of accumulation of evidence, $b_{d,s}$ the decision boundaries, and δ_s an offset representing time not directly related to the underlying evidence accumulation processes (e.g., the time required to encode the s th signal before evidence accumulation begins, etc.). We let $\theta_{d',s} = (\delta_s, \mu_{d',s}, b_{d',s})^T$.

Joint model for (d, τ) : Since a decision d is reached at response time τ if the corresponding threshold is crossed first, that is when $\{\tau = \tau_d\} \cap_{d' \neq d} \{\tau_{d'} > \tau_d\}$, we have $d = \arg \min_{d' \in \{1:d_0\}} \tau_{d'}$. Assuming simultaneous accumulation of evidence for all decision categories, modeled by independent Wiener processes, and termination when the threshold for the observed decision category d is reached, the joint distribution of (d, τ) is thus given by

$$f(d, \tau | s, \theta) = g(\tau | \theta_{d,s}) \prod_{d' \neq d} \{1 - G(\tau | \theta_{d',s})\}, \quad (2)$$

where, to distinguish from the generic notation f , we now use $g(\cdot | \theta)$ and $G(\cdot | \theta)$ to denote, respectively, the probability density function (pdf) and the cumulative distribution function (cdf) of an inverse Gaussian distribution, as defined in (1).

Marginal model for d : When the response times τ are unobserved, the probability of taking decision d given the stimulus s is thus obtained from (2) by integrating out the τ as

$$P(d | s, \theta) = \int_{\delta_s}^{\infty} g(\tau | \theta_{d,s}) \prod_{d' \neq d} \{1 - G(\tau | \theta_{d',s})\} d\tau. \quad (3)$$

The construction of model (3) is similar to traditional multinomial probit/logit regression models except that the latent variables are now inverse Gaussian distributed as opposed to being normal or extreme-value distributed, and the observed category is associated with the minimum of the latent variables in contrast to being identified with the maximum of the latent variables. We thus refer to this model as a ‘multinomial inverse-probit model’.

With data on both response categories d and response times τ available, the joint model (2) was used to construct the likelihood function in Paulon et al. (2021). In the absence of data on the response times τ , however, the inverse-probit model in (3) provides the basic building block for constructing the likelihood function for the observed response categories. As mentioned in the Abstract, discussed in the Introduction, and detailed in Sect. 2.1, the marginal inverse-probit model (3) for observed categories brings in many new identifiability issues and inference challenges not originally encountered for the joint model (2) developed in Paulon et al. (2021). Solving these new challenges for the marginal model (3) to infer the underlying drift-diffusion parameters $\theta_{d',s}$, for all d' , is the focus of this current article.

2.1. Identifiability Issues and Related Modeling Challenges

To begin with, we note that model (3) in itself cannot be identified from data on only the response categories. The offset parameters can easily be seen to not be identifiable since $P(\tau_d \leq \wedge_{d'} \tau_{d'}) = P\{(\tau_d - \delta) \leq \wedge_{d'} (\tau_{d'} - \delta)\}$ for any δ , where $\wedge_{d'} \tau_{d'}$ denotes the minimum of $\tau_{d'}$, $d' \in \{1 : d_0\}$. As is also well known in the literature, in categorical probability models, the location and scale of the latent continuous variables are not also separately identifiable. The following lemma establishes these points for the inverse-probit model.

Lemma 1. *The offset parameters δ_s are not identifiable in model (3). The drift and the boundary parameters, respectively $\mu_{d',s}$ and $b_{d',s}$, are also not separately identifiable in model (3).*

In the proof of Lemma 1 given in Appendix A, we have specifically shown that $P(d | s, \theta) = P(d | s, \theta^*)$, where the drift and boundary parameters in $\theta = \{(\mu_{d',s}, b_{d',s}); d' = 1, \dots, d_0\}$ and $\theta^* = \{(\mu_{d',s}^*, b_{d',s}^*); d' = 1, \dots, d_0\}$ satisfy $\mu_{d',s}^* = c\mu_{d',s}$ and $b_{d',s}^* = c^{-1}b_{d',s}$ for some constant $c > 0$. The result follows by noting that the transformation $\tau_{d',s}^* = c^{-2}\tau_{d',s}$ does not change the ordering between the $\tau_{d',s}$'s and hence the probabilities of the resulting decisions $d = \arg \min_{d' \in \{1:d_0\}} \tau_{d'} = \arg \min_{d' \in \{1:d_0\}} \tau_{d'}^*$ also remain the same. This has the simple implication that if the rate of accumulation of evidence is faster, then the same decision distribution is obtained if the corresponding boundaries are accordingly closer and conversely.

In fact, given the information on input and output categories alone, if d_0 denotes the number of possible decision categories, at most $d_0 - 1$ parameters are estimable. To see this, consider the probabilities $P(d' | s, \theta)$, $d' = 1, \dots, d_0$, where θ is the m -dimensional vector of parameters, possibly containing drift parameters and decision boundaries. Given the perceptual stimulus s as input, the probabilities satisfy $\sum_{d'=1}^{d_0} P(d' | s, \theta) = 1$. Thus, the function $P_s(\theta) = \{P(1 | s, \theta), \dots, P(d_0 | s, \theta)\}^T$ lie on a $d_0 - 1$ -dimensional simplex, $P_s(\theta) : \theta \rightarrow \Delta^{d_0-1}$, and by the model in (3) the mapping is continuous. Thus, it can be shown by the *Invariance of Domain* theorem (see, e.g., Deo, 2018) that if P_s is injective and continuous, then the domain of P_s must

belong to \mathbb{R}^m , where $m \leq d_0 - 1$. Thus in order to ensure identifiability of $\{P_s(\theta); \theta\}$, we must parametrize the probability vector with at most $d_0 - 1$ parameters.

The existing literature on drift-diffusion models discussed in the Introduction has traditionally put more emphasis on modeling the drifts (as their reference in the literature as ‘drift’-diffusion models suggests). Previous research on joint models for response times and associated response times in Paulon et al. (2021) also suggest that the boundaries remain stable around a value of 2 and it is primarily the changes in the drift rates that explain longitudinal learning. In view of this, we keep the boundaries fixed at the constant $b = 2$ and treat the drifts to be the free parameters instead. In our simulations and real data applications, it is observed that the estimates of the drift parameters and the associated cluster configurations are not very sensitive to small-to-moderate deviations of b around 2. In our codes implementing our method, available as part of the supplementary materials, we allow the practitioner to choose a value of b as they see fit for their specific application. The latent drift-diffusion process based with these constraints, namely $\delta_s = 0$ and $b_{d',s} = b$ for all d' , is shown in Fig. 1b.

While fixing $\delta_s = 0$ and $b_{d',s}$ to some known constant b reduces the size of the parameter space to d_0 , to ensure identifiability, we still need at least one more constraint on the drift parameters $\mu_{1:d_0,s}$. In categorical probability models, the identifiability problem of the location parameter is usually addressed by setting one category as a reference and modeling the probabilities of the others (Agresti, 2018; Albert & Chib, 1993; Borooah, 2002; Chib & Greenberg, 1998). However, posterior predictions from Bayesian categorical probability models with asymmetric constraints may be sensitive to the choice of reference category (see Burgette & Nordheim, 2012; Johndrow et al., 2013). Further, as also discussed in the Introduction, the goal of clustering the drift parameters $\mu_{1:d_0,s}$ across s can not be accomplished by this apparently simple solution.

The problem can be addressed by imposing a symmetric constraint on $\mu_{1:d_0,s}$ instead. A symmetric identifiability constraint has been previously proposed by Burgette et al. (2021) in the context of multinomial probit models, where they considered a sum-to-zero constraint on the latent utilities. To implement the constraint, they introduced a *faux base category* indicator parameter, which is assigned a discrete uniform prior and then learned via MCMC. Given this faux base category indicator, the other parameters are adjusted so that the sum-to-zero restriction is satisfied. However, the introduction of a base category, even if adaptively chosen, does not facilitate the clustering of $\mu_{1:d_0,s}$ within and across the different input categories s .

2.2. Our Proposed Approach

In coming up with solutions for these challenges, we take into consideration the complex design of our motivating tone learning experiments, so that our approach is easily extendable to longitudinal mixed model settings, allowing us to (a) estimate the smoothly varying trajectories of the parameters as the participants learn over time, (b) accommodate the heterogeneity between the participants, and (c) compare between the estimates not just within but also crucially between the different panels.

Similar to the *sum-to-zero* constraint in the multinomial probit model of Burgette et al. (2021), we impose a symmetric *sum to a constant* constraint on the drift parameters $\mu_{1:d_0,s}$ to identify our new class of inverse-probit models, although our implementation is quite different from theirs. To conduct inference, we start with an unconstrained prior, then sample from the corresponding unconstrained posterior, and finally project these samples to the constrained space through a minimal distance mapping. Similar ideas have previously been applied to satisfy natural constraints in other contexts. See, e.g., Dunson and Neelon (2003) and Gunn and Dunson (2005).

This approach is significantly advantageous both from a modeling and a computational perspective. On one hand, the basic building blocks are relatively easily extended to complex longitudinal mixed model settings, on the other, posterior computation is facilitated as this allows the

use of conjugate priors for the unconstrained parameters. Projection of the drift parameters onto the same space further makes them directly comparable, allowing clustering within and across the panels. The projected drifts can now be interpreted only on a relative scale but such compromises are not avoidable given the challenges we face.

2.2.1. Minimal Distance Mapping As the drift parameters are positive, the *sum to a constant* k constraint leads to constrained space $\mathcal{S}_k = \{\mu : \mathbf{1}^T \mu = k, \mu_j > 0, j = 1, \dots, d_0\}$ on which $\mu_{1:d_0,s}$ should be projected. The space \mathcal{S}_k is semi-closed, and therefore, the projection of any point μ onto \mathcal{S}_k may not exist. As a simple one dimensional example, let $x = -1$ and $\mathcal{S} = (0, 1]$, then $\arg \min_{y \in \mathcal{S}} |y - x| = 0 \notin \mathcal{S}$. Further, from a practical perspective, a drift parameter infinitesimally close to zero makes the distribution of the associated response times very flat which is typically not observed in real data. Therefore, we choose a small $\varepsilon > 0$ and project μ onto $\mathcal{S}_{\varepsilon,k} = \{\mu : \mathbf{1}^T \mu = k, \mu_j \geq \varepsilon, j = 1, \dots, d_0\}$. We then define the projection of a point μ onto $\mathcal{S}_{\varepsilon,k}$ through minimal distance mapping as

$$\mu^* = \text{Proj}_{\mathcal{S}_{\varepsilon,k}}(\mu) := \left\{ \arg \min_v \|\mu - v\| : v \in \mathcal{S}_{\varepsilon,k} \right\},$$

where $\|\cdot\|$ is the Euclidean norm. Note that for appropriate choices of (k, ε) , $\mathcal{S}_{\varepsilon,k}$ is non-empty, closed and convex. Therefore, μ^* exists and is unique by the Hilbert projection theorem (Rudin, 1991). The solution to this projection problem comes from the following result from Beck (2017).

Lemma 2. *Let $\mathcal{S}_{\varepsilon,k}$ be as defined above, and $\mathcal{S}_\varepsilon = \{\mu : \mu_j \geq \varepsilon, j = 1, \dots, d_0\}$. Then, $\text{Proj}_{\mathcal{S}_{\varepsilon,k}}(\mu) = \text{Proj}_{\mathcal{S}_\varepsilon}(\mu - u^* \mathbf{1})$, where u^* is a solution to the equation $\mathbf{1}^T \text{Proj}_{\mathcal{S}_\varepsilon}(\mu - u^* \mathbf{1}) = k$.*

Although the analytical form of the solution is not available, as is evident from the above result, the solution mainly relies on finding a root u^* of the non-increasing function $\phi(u^*) = \mathbf{1}^T \text{Proj}_{\mathcal{S}_\varepsilon}(\mu - u^* \mathbf{1}) - k$. We apply an algorithm based on Duchi et al. (2008) to reach the solution. The algorithm is described in ‘‘Appendix C’’.

2.2.2. Identifiability Restrictions The projection approach solves the problem of identifiability and maps the probability vector corresponding to an input tone s to the constraint space of $\mu_{1:d_0,s}$, $\mathcal{S}_{\varepsilon,k}$. The following theorem shows that the mapping from the constrained space of $\mu_{1:d_0,s}$ to the probability vector $\mathbf{P}(\mu_{1:d_0,s}) = \{p_1(\mu_{1:d_0,s}), \dots, p_{d_0}(\mu_{1:d_0,s})\}^T$ is injective. To keep the ideas simple, we consider the domain of the function to be $\mathcal{S}_{0,k}$ (i.e., $\varepsilon = 0$) instead of $\mathcal{S}_{\varepsilon,k}$ although a very similar proof would follow if $\mathcal{S}_{\varepsilon,k}$ were considered.

Theorem 1. *Let $p_d(\mu_{1:d_0,s})$ be the probability of observing the output tone d given the input tone s and the drift parameters $\mu_{1:d_0,s}$, as given in (3), for each $d = 1 : d_0$. Suppose $\mu_{1:d_0,s}$ lies on the space $\mathcal{S}_{0,k}$. Then, the function from $\mathcal{S}_{0,k}$ to the space of probabilities $\{p_d(\mu_{1:d_0,s}); d = 1 : d_0\}$ is injective.*

A proof is presented in ‘‘Appendix B’’.

2.2.3. Conjugate Priors for the Unconstrained Drifts From (3), given $\tau_1, \dots, \tau_{d_0}$, such that $\tau_d \leq \min\{\tau_1, \dots, \tau_{d_0}\}$, the posterior full conditional of $\mu_{1:d_0,s}$ is proportional to $\pi_\mu^{(n)} \propto \pi(\mu_{1:d_0,s}) \times \prod_{d'=1}^{d_0} g(\tau_{d'} | \mu_{d',s})$, where $\pi(\cdot)$ is the prior of $\mu_{1:d_0,s}$. Observe that $\prod_{d'=1}^{d_0} g(\tau_{d'} | \mu_{d',s})$ is Gaussian in $\mu_{1:d_0,s}$. A Gaussian prior on $\mu_{1:d_0,s}$ thus induces a conditional posterior for $\mu_{1:d_0,s}$ that is also Gaussian and hence very easy to sample from. Importantly, these benefits also extend naturally to multivariate Gaussian priors for any parameter vector $\beta_{d',s}$ that relates to $\mu_{d',s}$ linearly. This will be crucial in allowing us to extend the basic building block to longitudinal functional mixed model settings in Sect. 3 next, where we will be modeling time-varying $\mu_{d',s}(t)$ as flexible mixtures of B-splines with associated coefficients $\beta_{d',s}$.

2.2.4. *Justification as a Proper Bayesian Procedure* Define the constrained conditional posterior distribution, $\tilde{\pi}_{\mu}^{(n)}$, of the drift parameters μ as

$$\tilde{\pi}_{\mu}^{(n)}(B | \zeta) = \pi_{\mu}^{(n)}(\{\mu : \text{Proj}(\mu) \in B\} | \zeta), \quad B \subseteq \mathcal{S}_{\varepsilon,k},$$

where $\pi_{\mu}^{(n)}$ is the unconstrained conditional posterior of $\mu_{1:d_0,s}$, given the other variables ζ . The analytic form of the constrained conditional posterior is not available.

Sen et al. (2018) established a proper Bayesian justification for the posterior projection approach by showing the existence of a prior $\tilde{\pi}(\mu_{1:d_0,s})$ on the constrained space $\mathcal{S}_{\varepsilon,k}$ such that the resulting posterior is the same as the projected posterior $\tilde{\pi}_{\mu}^{(n)}$. When $\mathcal{S}_{\varepsilon,k}$ is non-empty, closed, and convex, i.e., the projection operator is measurable, such a prior exists if the unconstrained posterior is absolutely continuous with respect to the unconstrained prior (Sen et al., 2018, Corollary 1). As the unconstrained induced prior and posterior of the drift parameters are both Gaussian, this result holds in our case as well.

3. Extension to Longitudinal Mixed Models

In this section we adapt the inverse-probit model discussed in Sect. 2 to complex longitudinal design of our motivating PTC1 data set described in the Introduction. Let $s_{i,\ell,t}$ denote the input tone for the i th individual in the ℓ th trial of block t . Likewise, let $d_{i,\ell,t}$ denote, respectively, the output tone selected by the i th individual in the ℓ th trial of block t . Setting the offsets at zero, and boundary parameters to a fixed constant b , we now have

$$P\{d_{i,\ell,t} = d | s_{i,\ell,t} = s, \mu_{1:d_0,s}^{(i)}(t)\} = \int_0^{\infty} g\{\tau | \mu_{d,s}^{(i)}(t)\} \prod_{d' \neq d} \left[1 - G\{\tau | \mu_{d',s}^{(i)}(t)\}\right] d\tau,$$

$$\text{where } g\{\tau | s_{i,\ell,t} = s, \mu_{d',s}^{(i)}(t) = \mu\} = \frac{b}{\sqrt{2\pi}\tau^{3/2}} \exp\left[-\frac{\{b - \mu\tau\}^2}{2\tau}\right]. \quad (4)$$

The drift rates $\mu_{d',s}^{(i)}(t)$ now vary with the blocks t . In addition, we accommodate random effects by allowing $\mu_{d',s}^{(i)}(t)$ to also depend on the subject index i . We let $\mathbf{d} = \{d_{i,\ell,t}\}_{i,\ell,t}$, and d_0 be the number of possible decision categories (T1, T2, ..., T d_0). The likelihood function thus takes the form

$$L(\mathbf{d} | \mathbf{s}, \theta) = \prod_{d=1}^{d_0} \prod_{s=1}^{d_0} \prod_{t=1}^T \prod_{i=1}^n \prod_{\ell=1}^L \left[P\{d_{i,\ell,t} | s_{i,\ell,t}, \mu_{1:d_0,s}^{(i)}(t)\} \right]^{1\{d_{i,\ell,t}=d, s_{i,\ell,t}=s\}}.$$

We reiterate that in deriving the identifiability conditions and designing their implementation strategy in Sect. 2.2, we had to make sure that they would be applicable to the complex multi-subject longitudinal design of the PTC1 data set. Following those ideas, we model the time-varying mixed effects drift parameters $\mu_{d',s}^{(i)}(t)$ without any constraints first, then project them to the space satisfying the necessary identifying conditions.

For the unconstrained model, we follow the outline of Paulon et al. (2021) with necessary likelihood adjustments. The details are deferred to Section S.1 of the supplementary material. We present here a general outline.

We decompose $\mu_{d',s}^{(i)}(t) = f_{d',s}(t) + u_{d',s}^{(i)}(t)$ where $f_{d',s}(t)$ and $u_{d',s}^{(i)}(t)$ denote, respectively, fixed and random effects components, which are both modeled using flexible mixtures of B-spline bases. This allows us to cluster the fixed effects for different (d', s) combinations with similar shapes by clustering the corresponding B-spline coefficients.

Given posterior samples of $f_{d',s}(t)$ and $u_{d',s}^{(i)}(t)$, unconstrained samples of $\mu_{d',s}^{(i)}(t)$ are obtained. For every input tone s , these unconstrained $\mu_{1:d_0,s}^{(i)}(t)$'s are then projected to the space $\mathcal{S}_{\varepsilon,k}$ following the method described in Sect. 2.2.1.

4. Posterior Inference

Posterior inference for our proposed inverse-probit mixed model is carried out using samples drawn from the posterior using MCMC algorithm. The algorithm carefully exploits the conditional independence relationships encoded in the model as well as the latent variable construction of the model.

Inference can be greatly simplified by sampling the passage times $\tau_{1:d_0}$ and then conditioning on them. However, it is not possible to generate $\tau_{1:d_0}$ sequentially, e.g., by generating the passage time of the d -th decision choice τ_d independently, and that of the other decision choices from a truncated inverse-Gaussian distribution, left truncated at τ_d .¹

We implement a simple accept-reject sampler instead which generates values from the joint distribution of $\tau_{1:d_0}$ and accepts the sample if $\tau_d \leq \tau_{1:d_0}$. It is fast and produces a sample from the desired target conditional distribution. We formalize this result in the following lemma.

Lemma 3. *Let $g(\tau_{1:d_0} | \mu_{1:d_0})$ be the joint distribution of $\tau_{1:d_0}$. Consider the following accept-reject algorithm:*

Algorithm 1

Generating the passage times $\tau_{1:d_0}$ given $\arg \min_{d' \in \{1:d_0\}} \tau_{d'} = d$

- 1: Generate $\tau_{1:d_0}$ from the joint distribution $g(\tau_{1:d_0} | \mu_{1:d_0})$.
 - 2: Accept $\tau_{1:d_0}$ if $\tau_d \leq \tau_{1:d_0}$.
 - 3: Return to Step 1 otherwise.
-

Algorithm 1 generates samples from the conditional joint distribution of $\tau_{1:d_0}$, conditioned on the event $\tau_d \leq \tau_{1:d_0}$.

Proof of Lemma 3 is provided in ‘‘Appendix D’’.

It can be verified that the acceptance ratio of Algorithm 1 is $M^{-1} = P(\tau_d \leq \tau_{1:d_0})$ (see Robert & Casella, 2004) which depends on the drift parameters only. If the drift parameters are ordered accordingly, so as to satisfy $\mu_d \geq \mu_{1:d_0}$, the acceptance ratios increase. The algorithm thus becomes faster as the sampler converges.

As noted earlier, sampling the latent inverse-gaussian distributed response times $\tau_{1:d_0}$ greatly simplifies computation. Most of the chosen priors, including the priors on the coefficients β in

¹We can see this in a simpler example. Suppose we are interested in generating a sample from the conditional distribution of $\tau = (\tau_1, \tau_2)$ given $d = \arg \min_j \tau_j = 1$, where $\tau_i \sim \text{Uniform}(0, 1)$, $i = 1, 2$, independently. The conditional density of τ given $d = 1$ is $f_{\tau|d}(\tau_1, \tau_2) = 0.5$ if $0 < \tau_1 \leq \tau_2 < 1$, and = 0 otherwise. However, if we draw τ_1 from $\text{Uniform}(0, 1)$ first and let that realization be τ_1^* , and draw τ_2 from the truncated uniform distribution (left truncated at τ_1^*), then the pdf of the realization of (τ_1, τ_2) is τ_1^{*-1} .

the fixed and random effects, are conjugate. Due to space constraints, the details are deferred to Section S.3 in the supplementary material.

5. Simulation Studies

In this section, we discuss the results of a synthetic numerical experiment. We simulate data from a complex longitudinal design that mimics the real PTC1 data set. Our generating model contains fixed effects components attributed to different input-response tone combinations and random components attributed to individuals.

We recall that our main objective here is to identify the similarities and differences between the underlying brain mechanisms associated with different input-response category combinations over time while also assessing their individual heterogeneity, as characterized by latent drift-diffusion processes whose parameters can be biologically interpreted. The estimation of the probability curves for different input-response combinations, while a good indicator of our model's fit, is not the main purpose of this endeavor. Traditional categorical probability models, such as multinomial probit or logit, are thus not relevant to the scientific problem we are trying to address here. We are also not aware of any other work in the drift-diffusion literature that attempts to estimate the underlying parameters from category response data alone. In view of this, we restrict our focus to evaluating the performance of the proposed biologically meaningful longitudinal inverse-probit mixed model but do not present comparisons with any other model.

Design In designing the simulation scenario, we have tried to mimic our motivating category learning data sets. We chose $n = 20$ as the number of participants being trained over $T = 10$ blocks to identify $d_0 = 4$ tones. For each input tone and each block, there are $L = 40$ trials. We set the true $\mu_{d',s}(t)$ values in such a way that they are far from satisfying the constraint $\sum_{d'=1:d_0} \mu_{d',s} = k$, and the decision boundary is set to $b = 2$ for all (d', s) . The true drift parameters and the true probabilities, averaged over the participants of each input-response category combination, are shown in Fig. 3.

There are four true clusters in total, two for correct categorizations, S_1, S_2 , and two for incorrect categorizations, M_1, M_2 , as follows: $S_1 = \{(1, 1), (2, 2)\}$, $S_2 = \{(3, 3), (4, 4)\}$, $M_1 = \{(1, 2), (1, 3), (2, 1), (2, 3), (3, 4), (4, 3)\}$, $M_2 = \{(1, 4), (2, 4), (3, 1), (3, 2), (4, 1), (4, 2)\}$. We may interpret M_1 as the cluster of difficult alternatives, and M_2 as the cluster of easy alternatives. Thus, there are similarities in overall trajectories of $\{T_1, T_2\}$ and $\{T_3, T_4\}$, differentiating between easy and hard category recognition problems. We experimented with 50 synthetic data sets generated according to this design.

Results As the true drift parameters themselves do not satisfy the constraint, and the estimated drift parameters are on the constrained space, we cannot validate our method by its predictive performance of the drift parameters. Instead, the proposed method is validated in terms of the estimated probabilities.

Figure 4 shows the estimated posterior probability trajectories along with the 95% credible interval and the underlying true probability curves for every combination (d', s) in a typical scenario. The credible interval fails to capture the truth in two situations, when the true probability is very close to zero, or it is very close to one. The former case corresponds to classes with very low success probability, resulting in very few observations to estimate. The latter is underestimated as a consequence of the former since the probabilities add up to one.

The results produced by our method are mostly stable and consistent across all synthetic data sets. There are, however, a few cases of incorrect cluster assignments, resulting in some outliers in each boxplot. Note that if an incorrect cluster assignment takes place, the probabilities of all input-response combinations are affected by that. For example, if a component of M_1 is wrongly

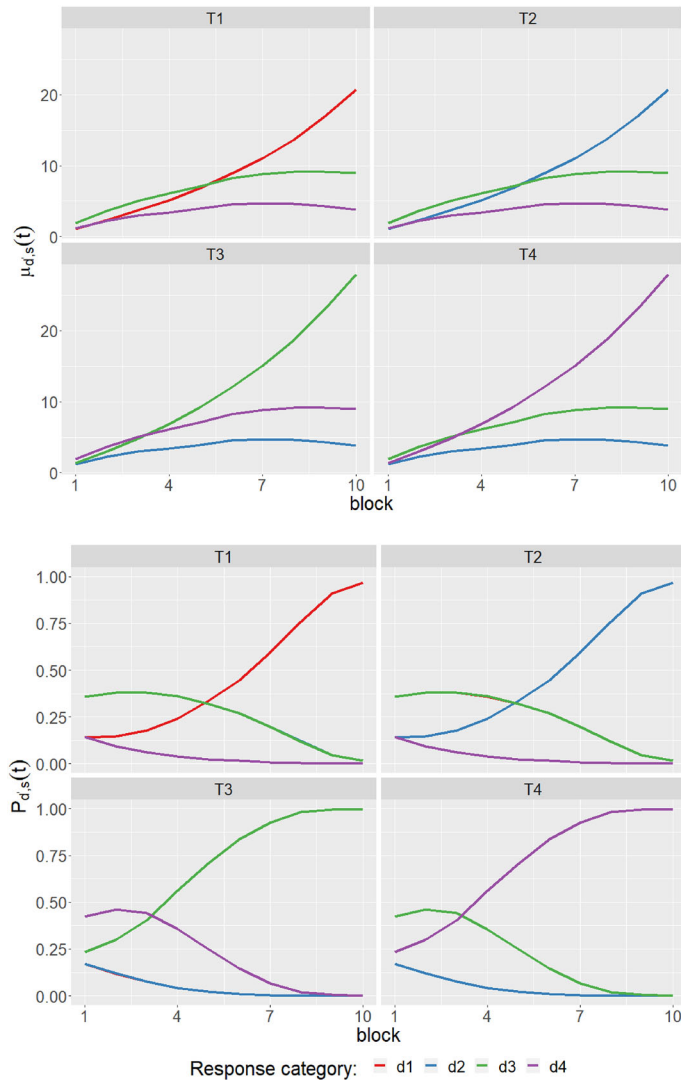


FIGURE 3.

Description of the synthetic data: True values of the drift parameters averaged over the subjects, denoted by $\mu_{d',s}(t)$, and true probabilities $P\{d_{i,\ell,t} | s_{i,\ell,t}, \mu_{1:d_{0,s}}^{(i)}(t)\}$ averaged over the subjects, denoted here by $P_{d',s}(t)$. Here T1, T2, T3, and T4 represent input categories 1 to 4, respectively. Some of the curves overlap according to the true clustering structure described in Sect. 5.

assigned to M_2 , then not only the probabilities of input–output combinations in M_1 and M_2 are affected, since the probabilities add up to one, those of S_1 and S_2 are also affected.

In estimating the probabilities, the overall mean squared error, i.e., the mean squared difference of the estimated and the true probabilities taking all combinations of (d, s, i, t) into account, came out to be 0.0028. Figure 5 provides a detailed description of the estimation of the probabilities for two input categories (one from each similarity group). As described for the individual simulation results, there are cases of under-estimation of the probabilities which are close to one, and consequently, over-estimation of the probabilities close to zero. However, the amount of

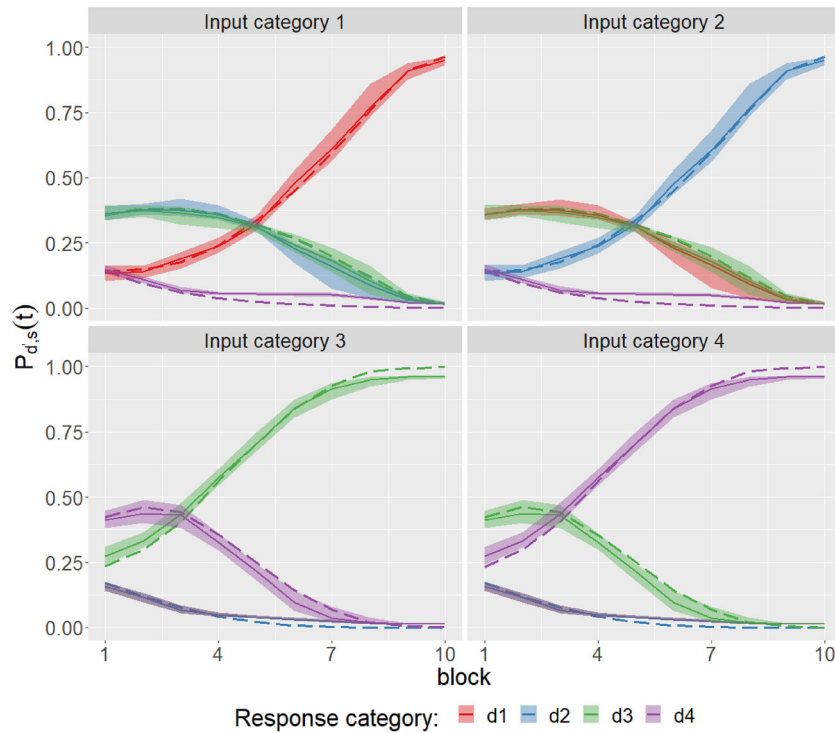


FIGURE 4.

Results for synthetic data: Posterior trajectories of the probabilities for each combination of (d', s) over blocks estimated by the proposed model. The shaded areas represent the corresponding 95% point-wise credible intervals. The thick dashed lines represent underlying true curves some of which overlap according to the true clustering structure described in Sect. 5.

departure from the true probability in each case is very small which can also be seen in the small overall MSE.

Further, the overall efficiency in identifying the true clustering structure is validated using Rand (Rand, 1971) and adjusted Rand (Hubert & Arabie, 1985) indices. The definitions of Rand and adjusted Rand indices are provided in Section S.6 in the supplementary material. The average Rand and adjusted Rand indices for our proposed method over 50 simulations 0.9105 and 0.8277, respectively, indicating high overall efficacy in correctly clustering the probability curves.

6. Applications

Analysis of the PTC1 Data Set We present here the analysis of the PTC1 data set described in Sect. 1 using our proposed longitudinal inverse-probit mixed model. We first demonstrate the performance of the proposed method in estimating the probabilities associated with different (d, s) pairs. Figure 6 shows the 95% credible intervals for the estimated probabilities for different input tones, along with the average proportions of times an input tone was classified into different tone categories across subjects. The latter serves as the empirical estimate of the probabilities.

We observe that except for the input-response combination (1, 1) in block 3 and some cases with a low number of data points, the 95% credible intervals include the corresponding empirical probabilities. An explanation of the occasional under-performance is given later in this section.

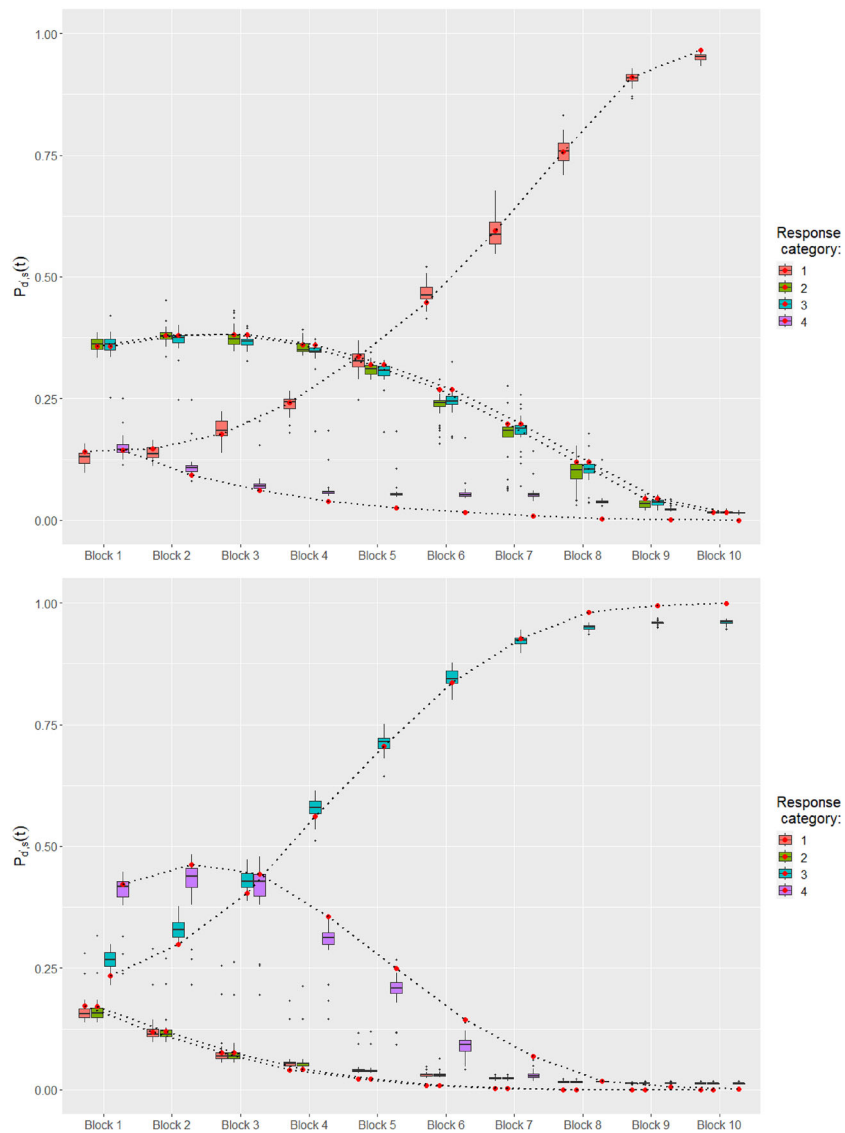


FIGURE 5.

Results of the synthetic data: Boxplots of the estimated probabilities over 50 simulations, and true probabilities (in red dot) of each block and for two panels, one from each similarity group (panel T_1 in the top and T_3 in the bottom).

Next, we examine the clusters identified by the proposed model. Apart from the two clusters obtained for the success combinations ($d = s$), three clusters are additionally identified in the incorrect input-response combinations ($d \neq s$). The clusters of success combinations are $S_1 = \{(1, 1), (2, 2), (4, 4)\}$ and $S_2 = \{(3, 3)\}$, and of wrong allocations are $M_1 = \{(1, 2), (1, 4), (2, 1), (2, 4), (3, 2), (4, 1), (4, 2)\}$, $M_2 = \{(1, 3), (2, 3), (3, 4), (4, 3)\}$, and $M_3 = \{(3, 1)\}$. Figure 7 shows the input-response tone combinations color-coded as per cluster identity, and the proportion of times each pair of input-response tone combinations appeared in the same cluster after burnin. Figure 7 indicates that, while the clusters S_1 , S_2 , M_1 are stable, there is some instability among the other two clusters, namely M_2 and M_3 .

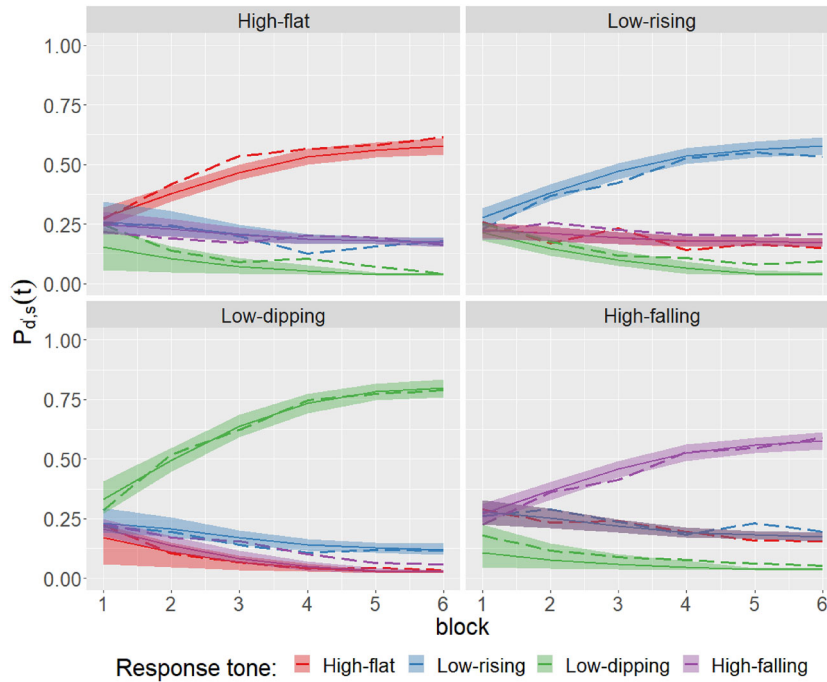


FIGURE 6.

Results for PTC1 data: Estimated probability trajectories compared with average proportions of times an input tone was classified into different tone categories across subjects (in dashed line). The means across subjects are indicated by thick lines and the shaded regions indicate corresponding 95% coverage regions.

Key Findings The clustering structure reveals that the low-dipping (T_3) response trajectories are different from the other three response categories. While for correct input–output tone combinations, S_2 forms a separate singleton cluster, for incorrect combinations, M_2 contains all the low-dipping trajectories, indicating their similarities across the panels. Also for T_3 , faster increase of the probabilities of correct identification, as well as faster decay of probabilities of incorrect identification indicate that T_3 is easily distinguishable from other alternatives.

On the other hand, the trajectories of high-flat (T_1), low-rising (T_2) and high-falling (T_4) response categories are quite similar across panels. While for correct input–response combinations, these three form the cluster S_1 , the corresponding incorrect tone combinations are clustered in M_1 . The slower rise of the observed empirical probabilities for the elements in S_1 and the slower decay of the same for M_1 indicate that T_1 , T_2 and T_4 are difficult to distinguish. However, in block 3 the empirical probabilities of correct input–response combinations differ moderately. While T_2 and T_4 show a relative drop in the empirical probabilities at block 3, T_1 shows a sudden pick in the same. This local dissimilarity of the trajectories at block 3, leads to a departure of the empirical probability of T_1 from the estimated credible band.

Next, we consider the results concerning the estimation of the drift parameters $\mu_{d',s}^{(i)}(t)$. As discussed in Sect. 2.2, given the identifiability constraints, the estimates of $\mu_{d',s}^{(i)}(t)$ can only be interpreted on a relative scale. Figure 8 shows the posterior mean trajectories and associated 95% credible intervals for the projected drift rates.

Importantly, our proposed mixed model also allows us to assess individual-specific parameter trajectories. Figure 9 shows the posterior mean trajectories and the associated 95% credible intervals for the drift rates $\mu_{d',s}^{(i)}$ estimated by our method for the different success combinations (d', s)

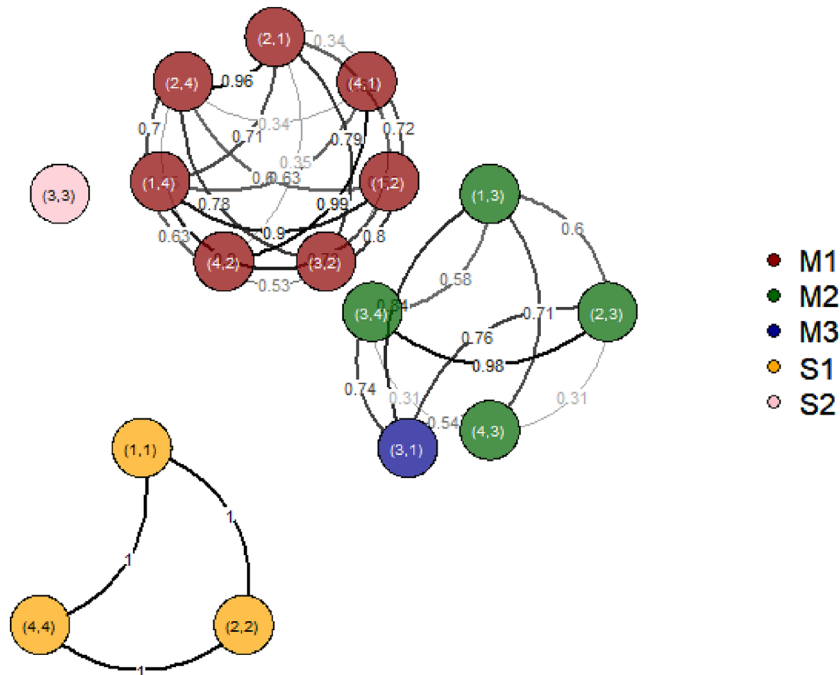


FIGURE 7.

Results for PTC1 data: Network plot of similarity groups showing the intra- and inter-cluster similarities of tone recognition problems. Each node is associated with a pair indicating the input-response tone category, (s, d) . The number associated with each edge indicates the proportion of times the pair in the two connecting nodes appeared in the same cluster after burnin.

for two participants—one with the best accuracy averaged across all blocks, and the other with the worst accuracy averaged across all blocks. These results suggest significant individual-specific heterogeneity. For the well-performing participant, the drift parameters are much higher than those for the poorly performing individual, indicating their ability to more quickly accumulate evidence compared to the poorly performing adult. These differences persisted over all blocks with a small gradual increase over time.

Analysis of Benchmark Data To validate the proposed method, we also analyzed tone learning data which, in addition to response accuracies, included accurate measurements of the response times. It was previously analyzed in Paulon et al. (2021) using the drift-diffusion model (2) which allowed inference on both the drift and the boundary parameters. For our analysis with the method proposed here, however, we ignored the response times. We observed that the estimates of the drifts produced by our proposed methodology match well with the estimates obtained by Paulon et al. (2021). A description of this ‘benchmark’ data set and other details of our analyses are provided in Section S.5 of the supplementary material.

7. Discussion, Conclusion, Broader Utility, and Future Work

Summary In this article, we developed a novel longitudinal inverse-probit mixed categorical probability model. Our research was motivated by category learning experiments where scientists are interested in using drift-diffusion models to understand how the decision-making mechanisms evolve as the participants get more training and experience. However, unlike traditional drift-

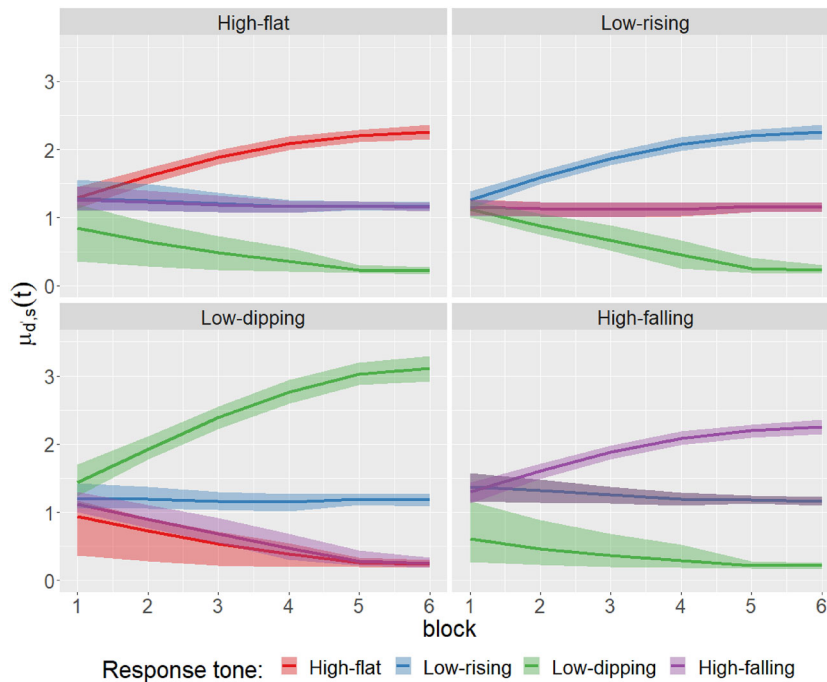


FIGURE 8.

Results for PTC1 data: Estimated posterior mean trajectories of the population level drifts $\mu_{d',s}(t)$ for the proposed model. The shaded areas represent the corresponding 95% point-wise credible intervals.

diffusion analyses which require data on both response categories and response times, we only had usable records of response categories but no response times. To our knowledge, biologically interpretable latent drift-diffusion process-based categorical probability models had never been considered for such scenarios in the literature before. We addressed this need. Building on a previous work on longitudinal drift-diffusion mixed joint models for response categories and response times but now integrating out the response times, we obtained a new class of category probability models which we referred to here as the inverse-probit model. We explored parameter recoverability in such models, showing, in particular, that the offset parameters can not be recovered and drifts and boundaries both can not be recovered from data only on response categories. In our analyses, we thus focused on estimating the biologically more important drift parameters but kept the offsets and the boundaries fixed. We showed that with careful domain knowledge informed choices for the boundaries, the general trajectories of the drift parameters can be recovered by our proposed approach even in the complete absence of response times.

Conclusion Overall, when it comes to making scientific inferences about drift-diffusion model parameters in the absence of data on response times, our work implies a mixed promise. On the downside, our work shows that the detailed interplay between drifts and boundaries cannot be captured. On the positive side, our results also suggest that, with our carefully designed model, and the fixed value of the boundary parameters appropriately chosen by experts, the general longitudinal trends in the drifts can still be estimated well from data only on response categories. Caution should still be exercised not to over-interpret the results.

Broader Utility in Auditory Neuroscience The proposed model, we believe, has significant implications for auditory neuroscience. We focused here specifically on a pupillometry study for which the experimental paradigms need to be adapted to prioritize slow pupillary response, rendering the

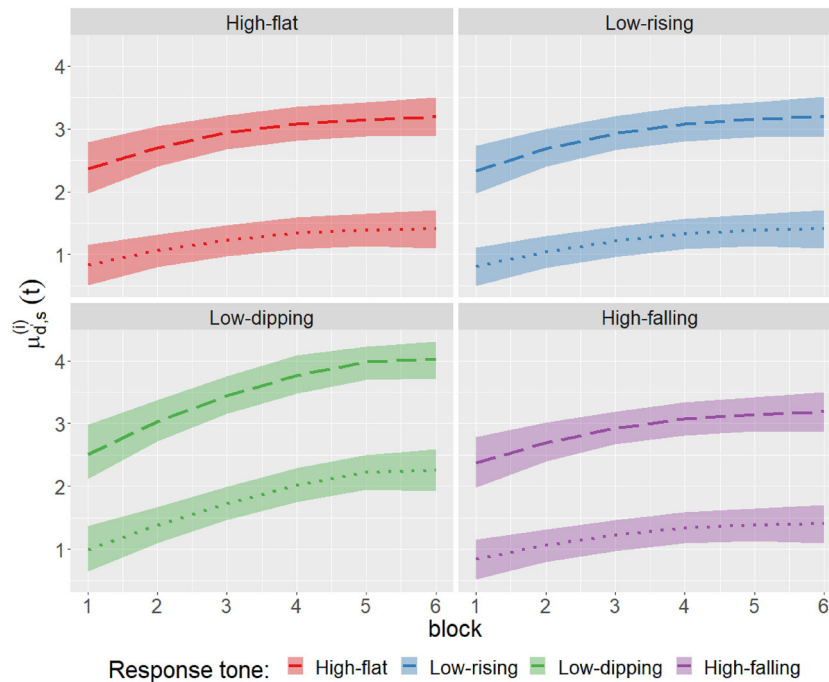


FIGURE 9.

Results for PTC1 data: Estimated posterior mean trajectories for individual specific drifts $\mu_{d',s}^{(i)}(t) = \exp\{f_{d',s}(t) + u_C^{(i)}(t)\}$ for successful identification ($d' = s$) for two different participants—one performing well (dashed line) and one performing poorly (dotted line). The shaded areas represent the corresponding 95% point-wise credible intervals.

behavioral response times useless. However, as discussed in the Introduction, there could be many other situations where usable data on response times may not be available. The proposed model can be useful in such scenarios to understand the perceptual mechanisms underlying auditory decision-making.

Broader Utility Beyond Auditory Neuroscience While we focused here on studying auditory category learning, the method proposed is applicable to other domains of behavioral neuroscience research studying categorical decision-making when the response times measurements are either not available or not reliable.

Broader Utility in Statistics On the statistical side, the projection-based approach proposed here to impose non-standard identifiability conditions and address clustering problems within and between different panels is not restricted to inverse-probit models introduced here. They can be easily adapted to other classes of generalized linear models such as the widely popular logit and probit models and hence may also be of interest to a much broader statistical audience.

Future Directions The models and the analyses of the PTC1 data set presented here excluded the pupillometry measurements themselves. An important and challenging problem being pursued separately elsewhere is to see how those measurements relate to drift-diffusion model parameters.

Funding This research was funded by the National Science Foundation grant DMS 1953712 and National Institute on Deafness and Other Communication Disorders Grants R01DC013315 and R01DC015504 awarded to Sarkar and Chandrasekaran.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Appendix

Appendix A: Proof of Lemma 1

Proof. It is easy to check that the offset parameters δ_s are not identifiable since

$$\begin{aligned} P(d | s, \delta_s, \mu_{1:d_0,s}, \mathbf{b}_{1:d_0,s}) &= \int_{\delta_s}^{\infty} g(\tau | \delta_s, \mu_{d,s}, b_{d,s}) \prod_{d' \neq d} \{1 - G(\tau | \delta_s, \mu_{d',s}, b_{d',s})\} d\tau \\ &= \int_0^{\infty} g(\tau | 0, \mu_{d,s}, b_{d,s}) \prod_{d' \neq d} \{1 - G(\tau | 0, \mu_{d',s}, b_{d',s})\} d\tau \\ &= P(d | s, 0, \mu_{1:d_0,s}, \mathbf{b}_{1:d_0,s}). \end{aligned}$$

Next we will show that the drift parameters and decision boundaries are not separately identifiable, even if we fix offset parameters to a constant.

First note that Eq. (3) can also be represented as

$$\int_{\delta_s}^{\infty} \dots \int_{\delta_s}^{\infty} \prod_{d' \neq d} g(\tau_{d'} | \theta_{d',s}) \int_{\delta_s}^{\wedge_{d' \neq d} \tau_{d'}} g(\tau_d | \theta_{d,s}) d\tau_d \prod_{d' \neq d} d\tau_{d'}. \quad (\text{A.1})$$

First observe that $\tau^* = \wedge_{d' \neq d} \tau_{d'} = \tau_{-1}^* \wedge \tau_1$, where $\tau_{-1}^* = \wedge_{d' \neq \{1,d\}} \tau_{d'}$. Thus the integral above can be written as

$$\begin{aligned} &\int_{\delta_s}^{\infty} \dots \int_{\delta_s}^{\infty} \prod_{d' \neq \{1,d\}} g(\tau_{d'} | \theta_{d',s}) \left\{ \int_{\delta_s}^{\infty} g(\tau_1 | \theta_{1,s}) \int_{\delta_s}^{\tau_{-1}^* \wedge \tau_1} g(\tau_d | \theta_{d,s}) d\tau_d \right\} \prod_{d' \neq \{1,d\}} d\tau_{d'} \\ &= \int_{\delta_s}^{\infty} \dots \int_{\delta_s}^{\infty} \prod_{d' \neq \{1,d\}} g(\tau_{d'} | \theta_{d',s}) \left\{ \int_{\delta_s}^{\tau_{-1}^*} g(\tau_d | \theta_{d,s}) \int_{\tau_d}^{\infty} g(\tau_1 | \theta_{1,s}) d\tau_1 d\tau_d \right\} \prod_{d' \neq \{1,d\}} d\tau_{d'}. \end{aligned}$$

Proceeding sequentially one can show that the integral above is the same as in (3).

Using the above we express the probability in (3) as in (A.1). As the offset parameter δ_s is already shown to be not identifiable, we need to fix the same. Without loss of generality, we fix the offset parameter at 0. The probability density function of inverse Gaussian distribution, with parameters $\theta_{d',s} = (\mu_{d',s}, b_{d',s})$ evaluated at $\tau_{d'}$, $g(\tau_{d'} | \theta_{d',s})$ can be obtained from (1) by replacing $\delta_s = 0$ and $d = d'$.

Consider the transformation of $\tau_{d'}$ to $\tau_{d'}^*$ as $\tau_{d'} = c^2 \tau_{d'}^*$, for some constant $c > 0$, and for all d' . Further, define $b_{d',s}^* = b_{d',s}/c$ and $\mu_{d',s}^* = c \mu_{d',s}$, for all d' . Then observe that

$$\begin{aligned} g(\tau_{d'} | \theta_{d',s}) d\tau_{d'} &= (2\pi)^{-1/2} b_{d',s}^* (\tau_{d'}^*)^{-3/2} \exp \left\{ -(2\tau_{d'}^*)^{-1} (b_{d',s}^* - \mu_{d',s}^* \tau_{d'}^*)^2 \right\} d\tau_{d'}^* \\ &= g(\tau_{d'}^* | \theta_{d',s}^*) d\tau_{d'}^*, \end{aligned}$$

where $g(\tau_{d'}^* | \theta_{d',s}^*)$ is the pdf of inverse Gaussian distribution with parameters $\mu_{d',s}^*$ and $b_{d',s}^*$, evaluated at the point $\tau_{d'}^*$.

Applying the transformation on $\tau_{d'}$ for all d' we get that the integral in (A.1) with $\delta_s = 0$ is same as

$$\int_0^\infty \cdots \int_0^\infty \prod_{d' \neq d} g(\tau_{d'}^* | \theta_{d',s}^*) \int_0^{\wedge_{d' \neq d} \tau_{d'}^*} g(\tau_d^* | \theta_{d,s}^*) d\tau_d^* \prod_{d' \neq d} d\tau_{d'}^*.$$

As c is arbitrary, this shows that the drifts and boundaries are not separately estimable. \square

Appendix B: Proof of Theorem 1

Proof. Let $\mathbf{P}(\mu_{1:d_0,s}) = \{p_1(\mu_{1:d_0,s}), \dots, p_{d_0}(\mu_{1:d_0,s})\}^T$ be the function, given by (4), from $\mathcal{S}_{0,k}$ to unit probability simplex Δ^{d_0-1} . For notational simplicity, we write $\mu_{1:d_0,s} = \mu = (\mu_1, \dots, \mu_{d_0})^T$. We first find the matrix of partial derivative $\nabla \mathbf{P}$ with respect to μ .

For $\mu \in \mathcal{S}_{0,k}$, $\mathbf{1}^T \mu = k$, and hence the probability reduces to

$$p_d(\mu) = \frac{(be^b)^{d_0}}{(2\pi)^{d_0/2}} \int_0^\infty \int_{\tau_d}^\infty \cdots \int_{\tau_d}^\infty |\tau|^{-3/2} \exp \left\{ -\frac{1}{2} (\mathbf{1}^T \tau^{-1} \mathbf{1} + \mu^T \tau \mu) \right\} d\tau_{-d} d\tau_d,$$

for $d = 1, \dots, d_0$, where $\tau = \text{diag}(\tau_1, \dots, \tau_{d_0})$, and τ_{-d} is the sub-vector of τ excluding the d -th element. Next, differentiating $p_d(\mu)$ with respect to μ , we get

$$\begin{aligned} \frac{\partial p_d(\mu)}{\partial \mu} &= \frac{(be^b)^{d_0}}{(2\pi)^{d_0/2}} \int_0^\infty \int_{\tau_d}^\infty \cdots \int_{\tau_d}^\infty |\tau|^{-3/2} (-\tau \mu) \exp \left\{ -\frac{1}{2} (\mathbf{1}^T \tau^{-1} \mathbf{1} + \mu^T \tau \mu) \right\} d\tau_{-d} d\tau_d, \\ &= [\mu_1 \eta_2 \quad \cdots \quad \mu_{d-1} \eta_2 \quad \mu_d \eta_1 \quad \mu_{d+1} \eta_2 \quad \cdots \quad \mu_{d_0} \eta_2]^T, \end{aligned}$$

where $\eta_1 = -E \{ \tau_1 \mathbb{I}(\tau_2 > \tau_1, \dots, \tau_{d_0} > \tau_1) | \mu \}$, and $\eta_2 = -E \{ \tau_2 \mathbb{I}(\tau_2 > \tau_1, \dots, \tau_{d_0} > \tau_1) | \mu \}$, and $\mathbb{I}(A)$ is the indicator function of the event A . Here the expectation is considered under the joint distribution of $(\tau_1, \dots, \tau_{d_0})$, which is independent inverse Gaussian. Clearly $\eta_1 > \eta_2 > 0$.

From the above derivation, it is easy to obtain that

$$\nabla \mathbf{P}(\mu) = \begin{bmatrix} \mu_1 \eta_1 & \mu_2 \eta_2 & \cdots & \mu_{d_0} \eta_2 \\ \mu_1 \eta_2 & \mu_2 \eta_1 & \cdots & \mu_{d_0} \eta_2 \\ \vdots & \vdots & \cdots & \vdots \\ \mu_1 \eta_2 & \mu_2 \eta_2 & \cdots & \mu_{d_0} \eta_1 \end{bmatrix} = \mathbf{M} \left\{ (\eta_1 - \eta_2) \mathbf{I} + \eta_2 \mathbf{1} \mathbf{1}^T \right\},$$

where $\mathbf{M} = \text{diag}(\mu_1, \dots, \mu_{d_0})$.

Now, suppose there exists μ and ν in \mathcal{S}_k such that $\mu \neq \nu$ and $\mathbf{P}(\mu) = \mathbf{P}(\nu)$. Define $\gamma : [0, 1] \rightarrow \mathbb{R}^{d_0}$ such that $\gamma(t) = \mu + t(\nu - \mu)$, $t \in [0, 1]$. Further, define $h(t) = \langle \mathbf{P}(\gamma(t)) - \mathbf{P}(\mu), \nu - \mu \rangle$, as the cross-product of $\mathbf{P}(\gamma(t)) - \mathbf{P}(\mu)$ and $\nu - \mu$. Then $h(1) = h(0) = 0$ under the proposition that $\mathbf{P}(\mu) = \mathbf{P}(\nu)$. Therefore, by the Mean Value Theorem, as $\mu \neq \nu$, there exists some point $c \in (0, 1)$ such that $\partial h(t)/\partial t|_{t=c} = 0$. Now,

$$\begin{aligned} \frac{\partial h(t)}{\partial t} &= \sum_{d'=1}^{d_0} (\nu_{d'} - \mu_{d'}) \frac{\partial}{\partial t} [p_{d'}\{\gamma(t)\} - p_{d'}(\mu)] \\ &= \sum_{d'=1}^{d_0} (\nu_{d'} - \mu_{d'}) \left\{ \frac{\partial}{\partial \gamma} p_{d'}(\gamma) \right\}^T \frac{\partial \gamma(t)}{\partial t} \\ &= (\nu - \mu)^T \nabla \mathbf{P}\{\gamma(t)\} (\nu - \mu) \\ &= (\eta_1 - \eta_2) (\nu - \mu)^T \Gamma(t) (\nu - \mu) + \eta_2 (\nu - \mu)^T \mathbf{M} \mathbf{1} \mathbf{1}^T (\nu - \mu) \\ &= (\eta_1 - \eta_2) (\nu - \mu)^T \Gamma(t) (\nu - \mu), \end{aligned}$$

as $\mathbf{1}^T (\nu - \mu) = 0$, where $\Gamma(t) = \text{diag}\{\gamma(t)\}$.

As every component of μ and ν is positive, for any $c \in (0, 1)$, the matrix $\Gamma(c)$ is positive definite. Further, as $\eta_1 > \eta_2$, $\partial h(t)/\partial t|_{t=c} = 0$ only if $\mu = \nu$, which contradicts the proposition. \square

Appendix C: Algorithm for Minimal Distance Mapping

The problem of finding projection of a point μ onto the space $\mathcal{S}_{k,\varepsilon}$ is equivalent to the following nonlinear optimization problem:

$$\text{minimize}_{\mathbf{w}} \|\mathbf{w} - \mu\|^2 \quad \text{such that} \quad \sum_{i=1}^{d_0} w_i = k, \quad w_i \geq \varepsilon.$$

Duchi et al. (2008, Algorithm 1) provides a solution to the problem of projection of a given point μ onto the space $\mathcal{S}_{k,\varepsilon}$ for $\varepsilon = 0$, which is modified for any given ε below.

INPUT: A vector μ , and a pair (k, ε) .

- 1: Sort μ into μ^* such that the elements of μ^* are in descending order.
- 2: Find $\rho = \max \left\{ j : \mu_j^* - j^{-1} \left(\sum_{l=1}^j \mu_l^* - k \right) > \varepsilon \right\}$.
- 3: Define $\theta = \rho^{-1} \left\{ \sum_{l=1}^{\rho} \mu_l^* - k + (k - \rho)\varepsilon \right\}$.

OUTPUT: \mathbf{w} such that $w_i = \max\{\mu_i - \theta, \varepsilon\}$.

Appendix D: Proof of Lemma 3

Proof. We consider the unconditional distribution of $\tau_{1:d_0}$, given the parameters $\mu_{1:d_0}$ as the proposal distribution, g . Clearly, the proposal distribution g and the target conditional joint distribution f satisfies $f(\tau_{1:d_0} | \mu_{1:d_0}) / g(\tau_{1:d_0} | \mu_{1:d_0}) \leq M$, where $M^{-1} = P(\tau_d \leq \tau_{1:d_0})$. Therefore, for any random sample $U \sim U(0, 1)$, $f(\tau_{1:d_0} | \mu_{1:d_0}) \geq MUg(\tau_{1:d_0} | \mu_{1:d_0})$ if the sample satisfies the condition $\tau_d \leq \tau_{1:d_0}$, and $f(\tau_{1:d_0} | \mu_{1:d_0}) < MUg(\tau_{1:d_0} | \mu_{1:d_0})$ otherwise. Hence, by Lemma 2.3.1 of Robert and Casella (2004), algorithm above produces samples from the target distribution. \square

References

- Agresti, A. (2018). *An introduction to categorical data analysis*. Wiley.
- Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, *88*, 669–679.
- Ashby, F. G., Noble, S., Filoteo, J. V., Waldron, E. M., & Ell, S. W. (2003). Category learning deficits in Parkinson's disease. *Neuropsychology*, *17*, 115.
- Beck, A. (2017). *First-order methods in optimization*. SIAM.
- Bogacz, R., Wagenmakers, E.-J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed-accuracy tradeoff. *Trends in Neurosciences*, *33*, 10–16.
- Borooh, V. K. (2002). *Logit and probit: Ordered and multinomial models*. Sage.
- Brody, C. D., & Hanks, T. D. (2016). Neural underpinnings of the evidence accumulator. *Current Opinion in Neurobiology*, *37*, 149–157.
- Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153–178.
- Burgette, L. F., & Nordheim, E. V. (2012). The trace restriction: An alternative identification strategy for the Bayesian multinomial probit model. *Journal of Business & Economic Statistics*, *30*, 404–410.
- Burgette, L. F., Puelz, D., & Hahn, P. R. (2021). A symmetric prior for multinomial probit models. *Bayesian Analysis*, *16*, 991–1008.
- Cavanagh, J. F., Wiecki, T. V., Cohen, M. X., Figueroa, C. M., Samanta, J., Sherman, S. J., & Frank, M. J. (2011). Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. *Nature Neuroscience*, *14*, 1462.
- Chandrasekaran, B., Yi, H.-G., & Maddox, W. T. (2014). Dual-learning systems during speech category learning. *Psychonomic Bulletin & Review*, *21*, 488–495.
- Chandrasekaran, B., Yi, H.-G., Smayda, K. E., & Maddox, W. T. (2016). Effect of explicit dimensional instruction on speech category learning. *Attention, Perception, & Psychophysics*, *78*, 566–582.
- Chhikara, R. (1988). *The inverse Gaussian distribution: Theory, methodology, and applications*. CRC Press.
- Chib, S., & Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, *85*, 347–361.
- Cox, D. R., & Miller, H. D. (1965). *The theory of stochastic processes*. CRC Press.
- de Boor, C. (1978). *A practical guide to splines*. Springer.
- Deo, S. (2018). *Algebraic topology*. Texts and Readings in Mathematics (Vol. 27). Hindustan Book Agency.
- Ding, L., & Gold, J. I. (2013). The basal ganglia's contributions to perceptual decision making. *Neuron*, *79*, 640–649.
- Duchi, J., Shalev-Shwartz, S., Singer, Y., & Chandra, T. (2008). Efficient projections onto the l_1 -ball for learning in high dimensions. In *Proceedings of the 25th international conference on machine learning* (pp. 272–279).
- Dufau, S., Grainger, J., & Ziegler, J. C. (2012). How to say “no” to a nonword: A leaky competing accumulator model of lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1117.
- Dunson, D. B., & Neelon, B. (2003). Bayesian inference on order-constrained parameters in generalized linear models. *Biometrics*, *59*, 286–295.
- Eilers, P. H., & Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical Science*, *11*, 89–102.
- Filoteo, J. V., Lauritzen, S., & Maddox, W. T. (2010). Removing the frontal lobes: The effects of engaging executive functions on perceptual category learning. *Psychological Science*, *21*, 415–423.
- Glimcher, P. W., & Fehr, E. (2013). *Neuroeconomics: Decision making and the brain*. Academic Press.
- Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574.
- Gunn, L. H., & Dunson, D. B. (2005). A transformation approach for incorporating monotone or unimodal constraints. *Biostatistics*, *6*, 434–449.
- Heekeren, H. R., Marrett, S., Bandettini, P. A., & Ungerleider, L. G. (2004). A general mechanism for perceptual decision-making in the human brain. *Nature*, *431*, 859.
- Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*, 193–218.
- Johndrow, J., Dunson, D., & Lum, K. (2013). Diagonal orthant multinomial probit models. In *Artificial intelligence and statistics* (pp. 29–38).
- Kim, S., Potter, K., Craigmile, P. F., Peruggia, M., & Van Zandt, T. (2017). A Bayesian race model for recognition memory. *Journal of the American Statistical Association*, *112*, 77–91.
- Lau, J. W., & Green, P. J. (2007). Bayesian model-based clustering procedures. *Journal of Computational and Graphical Statistics*, *16*, 526–558.
- Leite, F. P., & Ratcliff, R. (2010). Modeling reaction time and accuracy of multiple-alternative decisions. *Attention, Perception, & Psychophysics*, *72*, 246–273.
- Llanos, F., McHane, J. R., Schuerman, W. L., Yi, H. G., Leonard, M. K., & Chandrasekaran, B. (2020). Non-invasive peripheral nerve stimulation selectively enhances speech category learning in adults. *NPJ Science of Learning*, *1*, 1–11.
- Lu, J. (1995). Degradation processes and related reliability models. PhD thesis, McGill University, Montreal, Canada.
- McHane, J. R., Tessmer, R., Roark, C. L., & Chandrasekaran, B. (2021). Working memory relates to individual differences in speech category learning: Insights from computational modeling and pupillometry. *Brain and Language*, *22*, 1–15.
- Milosavljevic, M., Malmud, J., Huth, A., Koch, C., & Rangel, A. (2010). The drift diffusion model can account for the accuracy and reaction time of value-based choices under high and low time pressure. *Judgment and Decision Making*, *5*, 437–449.
- Morris, J. S. (2015). Functional regression. *Annual Review of Statistics and its Application*, *2*, 321–359.

- Parthasarathy, A., Hancock, K. E., Bennett, K., DeGruttola, V., & Polley, D. B. (2020). Bottom-up and top-down neural signatures of disordered multi-talker speech perception in adults with normal hearing. *Elife*, *9*, e51419.
- Paulon, G., Llanos, F., Chandrasekaran, B., & Sarkar, A. (2021). Bayesian semiparametric longitudinal drift-diffusion mixed models for tone learning in adults. *Journal of the American Statistical Association*, *116*, 1114–1127.
- Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing*, *39*, 204–214.
- Purcell, B. A. (2013). *Neural mechanisms of perceptual decision making*. Vanderbilt University.
- Ramsay, J. O., & Silverman, B. W. (2007). *Applied functional data analysis: Methods and case studies*. Springer.
- Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, *66*, 846–850.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59.
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*, 873–922.
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*, 347–356.
- Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*, 260–281.
- Reetzke, R., Xie, Z., Llanos, F., & Chandrasekaran, B. (2018). Tracing the trajectory of sensory plasticity across different stages of speech learning in adulthood. *Current Biology*, *28*, 1419–1427.
- Roark, C. L., Smayda, K. E., & Chandrasekaran, B. (2021). Auditory and visual category learning in musicians and nonmusicians. *Journal of Experimental Psychology: General*.
- Robert, C. P., & Casella, G. (2004). *Monte Carlo statistical methods*. Springer Texts in Statistics (2nd ed.). Springer.
- Robison, M. K., & Unsworth, N. (2019). Pupillometry tracks fluctuations in working memory performance. *Attention, Perception, & Psychophysics*, *81*, 407–419.
- Ross, S. M., Kelly, J. J., Sullivan, R. J., Perry, W. J., Mercer, D., Davis, R. M., Washburn, T. D., Sager, E. V., Boyce, J. B., & Bristow, V. L. (1996). *Stochastic processes*. Wiley.
- Rudin, W. (1991). *Functional analysis*. International Series in Pure and Applied Mathematics (2nd ed.). McGraw-Hill Inc.
- Schall, J. D. (2001). Neural basis of deciding, choosing and acting. *Nature Reviews Neuroscience*, *2*, 33.
- Sen, D., Patra, S., & Dunson, D. (2018). Constrained inference through posterior projections. arXiv preprint [arXiv:1812.05741](https://arxiv.org/abs/1812.05741).
- Smayda, K. E., Chandrasekaran, B., & Maddox, W. T. (2015). Enhanced cognitive and perceptual processing: A computational basis for the musician advantage in speech learning. *Frontiers in Psychology*, 1–14.
- Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*, 161–168.
- Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology*, *32*, 135–168.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*, 550.
- Wade, S. (2023). Bayesian cluster analysis. *Philosophical Transactions of the Royal Society A*, *381*, 1–20.
- Wang, J.-L., Chiou, J.-M., & Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and its Application*, *3*, 257–295.
- Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, *113*, 1033–1043.
- Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*, 3649–3658.
- Whitmore, G., & Seshadri, V. (1987). A heuristic derivation of the inverse gaussian distribution. *The American Statistician*, *41*, 280–281.
- Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started. *Trends in Hearing*, *22*, 1–32.
- Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear and Hearing*, *32*, 498–510.

Manuscript Received: 12 FEB 2023

Published Online Date: 19 FEB 2024