GLAMbox: A Python toolbox for investigating the association between gaze allocation and decision behaviour

Felix Molter*
WZB Berlin Social Science Center
Freie Universität Berlin

Armin W. Thomas*
Technische Universität Berlin

Hauke R. Heekeren
Freie Universität Berlin

Peter N. C. Mohr
WZB Berlin Social Science Center
Freie Universität Berlin

Abstract

Recent empirical findings have indicated that gaze allocation plays a crucial role in simple decision behaviour. Many of these findings point towards an influence of gaze allocation onto the speed of evidence accumulation in an accumulation-to-bound decision process (resulting in generally higher choice probabilities for items that have been looked at longer). Further, researchers have shown that the strength of the association between gaze and choice behaviour is highly variable between individuals, encouraging future work to study this association on the individual level. However, only few decision models exist that easily allow studying the gaze-choice association on the individual level, due to the high cost of developing and implementing them. The model space is particularly scarce for choice sets with more than two choice alternatives. Here, we present GLAMbox, a Python-based toolbox that is built upon PyMC3 and allows the easy application of the gaze-weighted linear accumulator model (GLAM) to experimental choice data. The GLAM assumes gaze-dependent evidence accumulation in a linear stochastic race that extends to decision scenarios with many choice alternatives. GLAMbox enables Bayesian parameter estimation of the GLAM for individual, pooled or hierarchical models, provides an easy-to-use interface to predict choice behaviour and visualize choice data, and benefits from all of PyMC3's Bayesian statistical modeling functionality. Further documentation, resources and the toolbox itself are available at https://github.com/glamlab/glambox.

Keywords: Decision making, visual attention, gaze bias, Python toolbox

A plethora of empirical findings has established an association between gaze allocation and decision behaviour on the group-level. For example, in value-based decision making, it was repeatedly shown that longer gaze towards one option is associated with a higher choice probability for that option (Armel, Beaumel, & Rangel, 2008; Cavanagh, Wiecki, Kochar, & Frank, 2014; Fiedler & Glöckner, 2012; Folke, Jacobsen, Fleming, & De Martino, 2017; Glöckner & Herbold, 2011; Konovalov & Krajbich, 2016; Krajbich, Armel, & Rangel, 2010; Krajbich & Rangel, 2011; Pärnamets et al., 2015; Shimojo, Simion, Shimojo, & Scheier, 2003; Stewart, Gächter, Noguchi, & Mullett, 2016; Stewart, Hermens, & Matthews, 2016; Vaidya & Fellows, 2015) and that external manipulation of gaze allocation changes choice probabilities accordingly (Armel et al., 2008; Pärnamets et al., 2015; Shimojo et al., 2003; Tavares, Perona, & Rangel, 2017). Such gaze bias effects are not limited to value-based decisions, but have recently also been observed in perceptual choices, where participants judge the perceptual attributes of stimuli based on available sensory information (Tavares et al., 2017).

These findings have led to the development of a set of computational models, aimed at capturing the empirically observed association between gaze allocation and choice behaviour by utilizing gaze data to inform the momentary accumulation rates of diffusion decision processes (Ashby, Jekel, Dickert, & Glöckner, 2016; Cavanagh et al., 2014; Fisher, 2017; Krajbich et al., 2010; Krajbich, Lu, Camerer, & Rangel, 2012; Krajbich & Rangel, 2011; Tavares et al., 2017). Specifically, these models assume that evidence accumulation in favour of an item continues while it is not looked at, but at a discounted rate. The application of these models is limited so far, as fitting them to empirical data depends on computationally expensive simulations, involving the simulation of fixation trajectories. These simulations, as well as the creation of models of the underlying fixation process, become increasingly difficult with increasing complexity of the decision setting (e.g., growing choice set sizes or number of option attributes, etc.). Existing approaches that circumvent the need for simulations, model the evidence accumulation process as a single diffusion process between two decision bounds and are therefore limited to binary decisions (Cavanagh et al., 2014; Lopez-Persem, Domenech, & Pessiglione, 2016).

However, researchers are increasingly interested in choice settings involving more than two alternatives. Choices outside the laboratory usually involve larger choice sets or describe items on multiple attributes. Besides, many established behavioural effects only occur in multi-alternative and multi-attribute choice situations (e.g., context effects; Trueblood, Brown, Heathcote, & Busemeyer, 2013).

Furthermore, recent findings indicate strong individual differences in the association

---

Felix Molter and Peter N. C. Mohr, Freie Universität Berlin, School of Business & Economics, Garystr. 21, 14195 Berlin, Germany; WZB Berlin Social Science Center, Reichpietschufer 50, 10785 Berlin, Germany. Armin W. Thomas, Technische Universität Berlin, Marchstr. 23, 10587 Berlin, Germany; Max Planck School of Cognition, Stephanstr. 1A, 04103 Leipzig, Germany. Hauke R. Heekeren, Freie Universität Berlin, Department of Education and Psychology, Habelschwerdter Allee 45, 14195 Berlin, Germany

* These authors contributed equally to the work.

Correspondence should be addressed to Felix Molter (felixmolter@gmail.com) or Armin W. Thomas (athms.research@gmail.com)

between gaze allocation and choice behaviour (Smith & Krajbich, 2018; Thomas, Molter, Krajbich, Heekeren, & Mohr, 2019) as well as individual differences in the decision mechanisms used (Ashby et al., 2016). Yet, the majority of model-based investigations of the relationship between gaze allocation and choice behaviour were focused on the group level, disregarding differences between individuals.

With the Gaze-weighted linear accumulator model (GLAM; Thomas et al., 2019), we have proposed an analytical tool that allows the model-based investigation of the relationship between gaze allocation and choice behaviour at the level of the individual, in choice situations involving more than two alternatives, solely requiring participants' choice, response time (RT) and gaze data, in addition to estimates of the items' values.

The GLAM assumes that the decision process is biased by momentary gaze behaviour. While an item is not fixated, its value representation is discounted. The resulting decision signals are averaged using relative gaze data, compared, transformed and fed into a stochastic race framework. Thereby, GLAM naturally generalizes to choice scenarios involving more than two alternatives while remaining analytically tractable. By providing an analytical solution for the first-passage-density (FPD) of the stochastic race process, these models' parameters can be efficiently estimated in a hierarchical Bayesian framework.

To make GLAM more accessible, we now introduce GLAMbox, a Python-based toolbox for the application of the GLAM to empirical choice, RT and gaze data. GLAMbox allows for individual and hierarchical estimation of the GLAM parameters, simulation of response data and model-based comparisons between experimental conditions and groups. It further contains a set of visualization functions to inspect choice and gaze data and evaluate model fit. We illustrate three application examples of the toolbox: In Example 1, we illustrate how GLAMbox can be used to analyze individual participant data with the GLAM. In particular, we perform an exemplary model comparison between multiple model variants on the individual level, as well as an out-of-sample prediction of participants' choice and RT data. In Example 2, we demonstrate the application of the GLAM to perform a comparison of group-level parameters in a setting with limited amounts of data, using hierarchical parameter estimation. Lastly, in Example 3, we walk the reader through a step-by-step parameter recovery study with the GLAM, which is encouraged to increase confidence in the estimated parameter values.

Methods

Gaze-weighted linear accumulator model details

The GLAM assumes that preference formation, during a simple choice process, is guided by the allocation of visual gaze (for an overview, see Fig. 1). Particularly, the decision process is guided by a set of decision signals: An absolute and relative decision signal. Throughout the trial, the absolute signal of an item can be in two states: An unbiased state, equal to the item's value r_i while the item is looked at, and a biased state while any other item is looked at, where the item value r_i is discounted by a parameter γ. The average absolute decision signal $\bar{A}_i$ is given by

$$\bar{A}_i = g_i r_i + (1 - g_i)\gamma r_i, \tag{1}$$

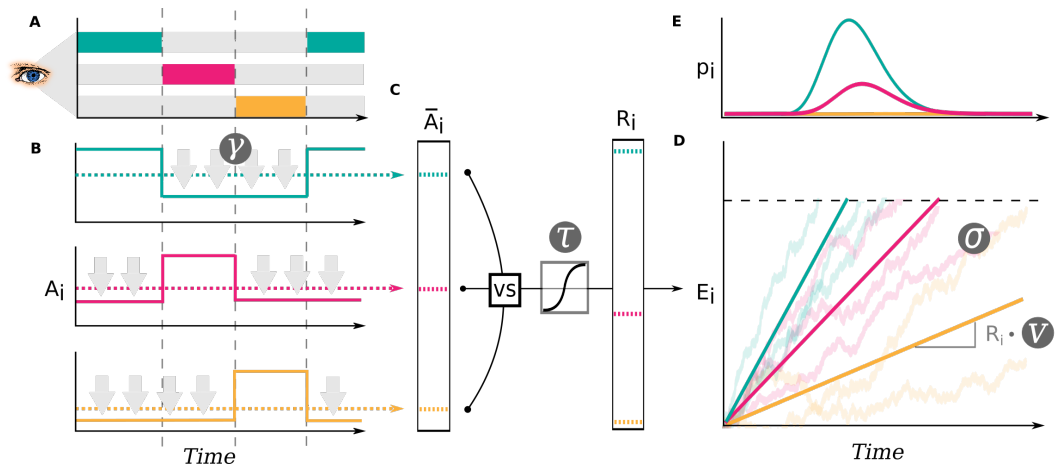

Figure 1. In the GLAM, preference formation during the decision process is dependent on the allocation of visual gaze (A). For each item in the choice set, an average absolute decision signal $\bar{A}_i$ is computed (dashed lines in B). The magnitude of this signal is determined by the momentary allocation of visual gaze: While an item is currently not looked at, its signal is discounted by parameter γ ($\gamma \leq 1$; discounting is illustrated by gray arrows) (B). To determine a relative decision signal R_i for each item in the choice set, absolute evidence signals are transformed in two steps (C): First, the difference between each average absolute decision signal $\bar{A}_i$ and the maximum of all others is determined. Second, the resulting differences are scaled through a logistic transform, as the GLAM assumes an adaptive representation of the relative decision signals that is especially sensitive to differences close to 0 (where the absolute signal for an item is very close to the maximum of all others). The resulting relative decision signals R_i can be used to predict choice and RT, by determining the speed of the accumulation process in a linear stochastic race (D). The stochastic race then provides first-passage time distributions p_i, describing the likelihood of each item being chosen at each time point (E).

where g_i is defined as the fraction of total trial time that item i was looked at. If $\gamma = 1$, there is no difference between the biased and unbiased state, resulting in no influence of gaze allocation on choice behaviour. For γ values less than 1, the absolute decision signal A_i is discounted, resulting in generally higher choice probabilities for items that have been looked at longer.

To determine the relative decision signals, the average absolute decision signals $\bar{A}_i$ are transformed in two steps: First, The difference R_i^* between the average absolute decision signal $\bar{A}_i$ and the maximum of all other decision signals $\bar{A}_J$ is computed:

$$R_i^* = \bar{A}_i - \max_J \bar{A}_J. \tag{2}$$

Second, the resulting difference signals R_i^* are scaled through a logistic transform $s(x)$. The GLAM assumes an adaptive representation of the relative decision signals, which is maximally sensitive to small differences in the absolute decision signals close to 0 (where the difference between the absolute decision signal of an item and the maximum of all others

is small):

$$R_i = s(R_i^*) \tag{3}$$

$$s(x) = \frac{1}{1 + \exp -\tau x} \tag{4}$$

The sensitivity of this transform is determined by the temperature parameter τ of the logistic function. Larger values of τ indicate stronger sensitivity to small differences in the average absolute decision signals $\bar{A}_i$.

To lastly capture response behaviour as well as RTs, the relative signals R_i can be fed into a linear stochastic race. Here, one item accumulator E_i is defined for each item in the choice set:

$$E_i(t) = E_i(t-1) + vR_i + N(0, \sigma^2), \text{with } E_i(0) = 0 \tag{5}$$

At each time step t, the amount of accumulated evidence is determined by the accumulation rate vR_i, (where v indicates a general speed parameter that is independent of gaze and item value) and zero-centered normally distributed noise with standard deviation σ. A choice for an item is made as soon as one accumulator reaches the decision boundary b. The first passage time density $f_i(t)$ of a single linear stochastic accumulator E_i, with decision boundary b, is given by the inverse Gaussian distribution:

$$f_i(t) = \Big[\frac{\lambda}{2\pi t^3}\Big]^{\frac{1}{2}} \exp\Big(\frac{-\lambda(t-\mu)^2}{2\mu^2 t}\Big) \tag{6}$$

$$\text{with } \mu = \frac{b}{vR_i} \text{ and } \lambda = \frac{b^2}{\sigma^2}$$

However, this density does not take into account that there are multiple accumulators in each trial racing towards the same boundary. For this reason, $f_i(t)$ must be corrected for the probability that any other accumulator crosses the boundary first. The probability that an accumulator crosses the boundary prior to t, is given by its cumulative distribution function $F_i(t)$:

$$F_i(t) = \Phi\Big(\sqrt{\frac{\lambda}{t}}(\frac{t}{\mu} - 1)\Big) + \exp\Big(\frac{2\lambda}{\mu}\Big) \cdot \Phi\Big(-\sqrt{\frac{\lambda}{t}}(\frac{t}{\mu} + 1)\Big) \tag{7}$$

Here, $\Phi(x)$ defines the standard normal cumulative distribution function. Hence, the joint probability $p_i(t)$ that accumulator E_i crosses b at time t, and that no other accumulator E_j has reached b first, is given by:

$$p_i(t) = f_i(t) \prod_J (1 - F_j(t)) \tag{8}$$

Contaminant response model. To reduce the influence of erroneous responses (e.g., when the participant presses a button by accident or has a lapse of attention during the task) on parameter estimation, we include a model of contaminant response processes in all estimation procedures: In line with existing drift diffusion modelling toolboxes (e.g., Wiecki, Sofer, & Frank, 2013), we assume a fixed 5% rate of erroneous responses ϵ that is modeled as a participant-specific uniform likelihood distribution $u_s(t)$. This likelihood describes the probability of a random choice for any of the N available choice items at a random time point in the interval of empirically observed RTs (cf., Ratcliff & Tuerlinckx, 2002; Wiecki et al., 2013):

$$u_s(t) = \frac{1}{N(\max \mathrm{rt}_s - \min \mathrm{rt}_s)} \tag{9}$$

The resulting likelihood for participant s choosing item i, accounting for erroneous responses, is then given by:

$$l_i(t) = (1 - \epsilon) \cdot p_i(t) + \epsilon \cdot u_s(t) \tag{10}$$

The rate of error responses ϵ can be specified by the user to a different value than the default of 5% using the `error_weight` keyword in the `make_model` method (see below).

Individual parameter estimation details. The GLAM is implemented in a Bayesian framework using the Python library PyMC3 (Salvatier, Wiecki, & Fonnesbeck, 2016, version 3.7). The model has four parameters (v, γ, σ, τ). By default, uninformative, uniform priors between sensible limits (derived from earlier applications to four different datasets; Thomas et al., 2019) are placed on all parameters:

$$\begin{aligned} v &\sim U(0, 2) \\ \gamma &\sim U(-10, 1) \\ \sigma &\sim U(0, 2) \\ \tau &\sim U(0, 5) \end{aligned}$$

The γ parameter has a natural upper bound at 1 (indicating no gaze bias). The τ parameter has a natural lower bound at 0 (resulting in no sensitivity to differences in average absolute decision signals $\bar{A}_i$). The velocity parameter v and the noise parameter σ must be strictly positive.

Hierarchical parameter estimation details. For hierarchical models, individual parameters are assumed to be drawn from Truncated Normal distributions, parameterized by mean and standard deviation, over which weakly informative, Truncated Normal priors are assumed (based on the distribution of group level parameter estimates obtained from four different datasets in Thomas et al. 2019; see Figs. 2, A2 and A3 and Table A1):

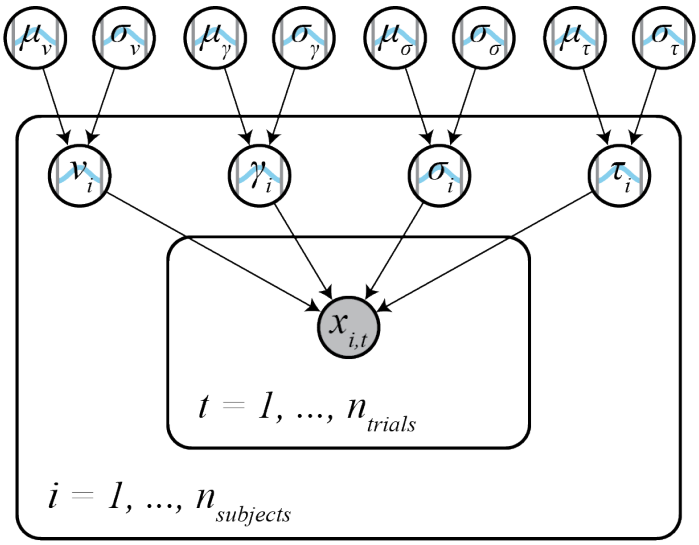

Figure 2. In the hierarchical model, individual subject parameters γ_i, v_i, σ_i and τ_i (subject plate) are assumed to be drawn from Truncated Normal group level distributions with means μ and standard deviations σ (outside of subject plate). By default, weakly informative Truncated Normal priors, based on previous on previous applications of the model to four different datasets, are placed on the group level parameters. RT and choice data $x_{i,t}$ for each trial t is distributed according to subject parameters and the GLAM likelihood (Eq. 8; inner trial plate).

$$
\begin{aligned}
v_\mu &\sim N(0.63, 10 \cdot 0.26), \text{truncated to } [0, 2] \\
v_\sigma &\sim N(0.26, 10 \cdot 0.11), \text{truncated to } [0, 1] \\
\gamma_\mu &\sim N(0.12, 10 \cdot 0.11), \text{truncated to } [-2, 1] \\
\gamma_\sigma &\sim N(0.35, 10 \cdot 0.1), \text{truncated to } [0, 1] \\
\sigma_\mu &\sim N(0.27, 10 \cdot 0.08), \text{truncated to } [0, 1] \\
\sigma_\sigma &\sim N(0.05, 10 \cdot 0.01), \text{truncated to } [0, 0.2] \\
\tau_\mu &\sim N(1.03, 10 \cdot 0.58), \text{truncated to } [0, 5] \\
\tau_\sigma &\sim N(0.62, 10 \cdot 0.26), \text{truncated to } [0, 3]
\end{aligned}
$$

Basic usage

Data format, the `GLAM` class. The core functionality of the GLAMbox is implemented in the `GLAM` model class. To apply the GLAM to data, an instance of the model class needs to be instantiated and supplied with the experimental data, first:

```
import glambox as gb
glam = gb.GLAM(data=data)
```

subject	trial	choice	rt	item_value_0	item_value_1	item_value_2	gaze_0	gaze_1	gaze_2	speed
0	0	0	2.056	5	1	3	0.16	0.62	0.22	'fast'
0	1	2	3.685	3	6	9	0.44	0.22	0.34	'slow'

Table 1
The first two rows of a pandas DataFrame ready to be used with GLAM.

The data must be a pandas (McKinney, 2010) DataFrame with one row per trial, containing the following variable entries:

- `subject`: Subject index (integer, first subject should be 0)
- `trial`: Trial index (integer, first trial should be 0)
- `choice`: Chosen item in this trial (integer, items should be 0, 1, $\ldots, N$)
- `rt`: Response time (float, in seconds)
- for each item i in the choice set:
  - `item_value_i`: The item value (float)
  - `gaze_i`: The fraction of total time in this trial that the participant spent looking at this item (float, between 0 and 1)
- additional variables coding groups or conditions (string or integer)

For reference, the first two rows of a pandas DataFrame ready to be used with GLAMbox are shown in Table 1.

Next, the respective PyMC3 model, which will later be used to estimate the model's parameters, can be built using the `make_model` method. Here, the researcher specifies the kind of the model: `'individual'` if the parameters should be estimated for each subject individually, `'hierarchical'` for hierarchical parameter estimation, or `'pooled'` to estimate a single parameter set for all subjects. At this stage, the researcher can also specify experimental parameter dependencies: For example, a parameter could be expected to vary between groups or conditions. In line with existing modeling toolboxes (e.g., Voss & Voss, 2007; Wiecki et al., 2013) dependencies are defined using the `depends_on` argument. `depends_on` expects a dictionary with parameters as keys and experimental factors as values (e.g., `depends_on=dict(v='speed')` for factor `'speed'` with conditions `'fast'` and `'slow'` in the data). The toolbox internally handles within- and between subject designs and assigns parameters accordingly. If multiple conditions are given for a factor, one parameter will be designated for each condition. Finally, the `make_model` method allows parameters to be fixed to a specific value using the `*_val` arguments (e.g., `gamma_val=1` for a model without gaze bias). If parameters should be fixed for individual subjects, a list of individual values needs to be passed.

```
model.make_model(kind='individual',
                 depends_on=dict(v='speed'),
                 gamma_val=1)
```

Inference. Once the PyMC3 model is built, parameters can be estimated using the `fit` method:

```
model.fit(method='MCMC')
```

The `fit` method defaults to Markov-Chain-Monte-Carlo (MCMC; Gamerman & Lopes, 2006) sampling, but also allows for Variational Inference (see below).

Markov-Chain-Monte-Carlo. MCMC methods approximate the Bayesian posterior parameter distributions, describing the probability of a parameter taking certain values given the data and prior probabilities, through repeated sampling. GLAMbox can utilize all available MCMC step methods provided by PyMC3. The resulting MCMC traces can be accessed using the `trace` attribute of the model instance (note that a list of traces is stored for models of kind `'individual'`). They should always be checked for convergence, to ascertain that the posterior distribution is approximated well. Both qualitative visual and more quantitative numerical checks of convergence, such as the Gelman-Rubin statistic $\hat{R}$ and the number of effective samples are recommended (see Gelman & Shirley, 2011; Kruschke, 2014, for detailed recommendations). PyMC3 contains a range of diagnostic tools to perform such checks (such as the `summary` function).

Variational inference. Estimation can also be done using all other estimation procedures provided in the PyMC3 library. This includes variational methods like Automatic Differentiation Variational Inference (ADVI; Kucukelbir, Tran, Ranganath, Gelman, & Blei, 2017). To use variational inference, the `method` argument can be set to `'VI'`, defaulting to the default variational method in PyMC3. We found variational methods to quickly yield usable, but sometimes inaccurate parameter estimates, and therefore recommend using MCMC for final analyses.

Accessing parameter estimates. After parameter estimation is completed, the resulting estimates can be accessed with the `estimates` attribute of the GLAM model instance. This returns a table with one row for each set of parameter estimates for each individual and condition in the data. For each parameter, a maximum a posteriori (MAP) estimate is given, in addition to the 95% Highest-Posterior Density Interval (HPD). If the parameters were estimated hierarchically, the table also contains estimates of the group-level parameters.

Predicting choices and response times. Choices and RTs can be predicted with the GLAM by the use of the `predict` method:

```
model.predict(n_repeats=50)
```

For each trial of the dataset that is attached to the model instance, this method predicts a choice and RT according to Eq. 10, using the previously determined *maximum a posteriori* (MAP) parameter estimates. To obtain a stable estimate of the GLAM's predictions, as well as the noise contained within them, it is recommended to repeat every trial multiple times during the prediction. The number of trial repeats can be specified with the `n_repeats` argument. After the prediction is completed, the predicted data can be accessed with the `prediction` attribute of the model.

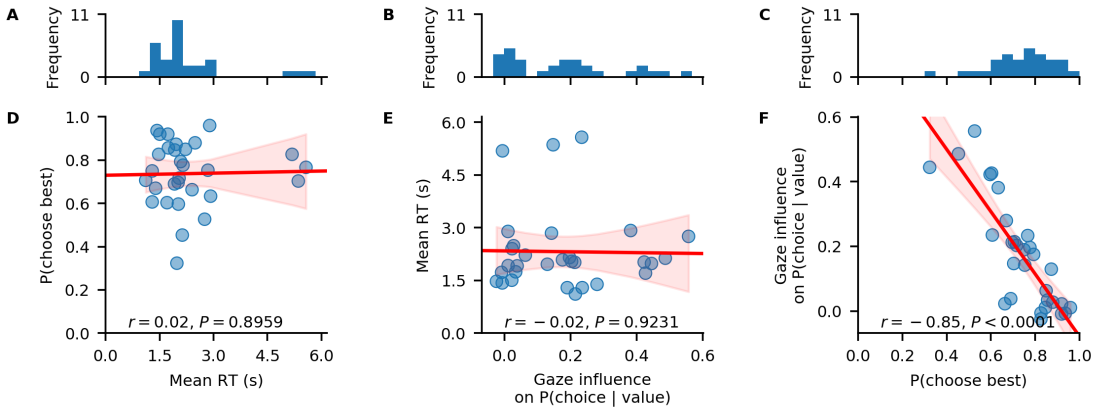

Figure 3. Individual differences in the data. A-C: distributions of individuals' mean RT (A), behavioural influence of gaze allocation on choice behaviour (B), and the probability of choosing the highest-valued item in a trial (C). D-F: associations between individuals' mean RT and their probability of choosing the highest-valued item (D), individuals' behavioural influence of gaze allocation on choice behaviour and their mean RT (E), individuals' behavioural influence of gaze allocation on choice behaviour and their probability of choosing the highest-valued item (F). Red lines indicate linear regression fits with confidence bands surrounding them. Pearson's r coefficients with corresponding P-values are reported for each association in D-F.

Results

Example 1: Individual Level Data & Model Comparison

Our first example is based on the study by Thomas et al. (2019). Here, the authors study the association between gaze allocation and choice behaviour on the level of the individual. In particular, they explore whether (1) gaze biases are present on the individual level and (2) the strength of this association varies between individuals. In this example, we replicate this type of individual model-based analysis, including parameter estimation, comparison between multiple model variants, and out-of-sample prediction of choice and RT data.

Simulating data. First, we simulate a dataset containing 30 subjects, each performing 300 simple value-based choice trials. We assume that in each trial participants are asked to choose the item that they like most out of a set of three presented alternatives (e.g., snack food items; similar to the task described in Krajbich and Rangel (2011)). While participants perform the task, their eye movements, choices and RTs are measured. After completing the choice trials, participants are further asked to indicate their liking rating for each of the items used in the choice task on a liking rating scale between 1 and 10 (with 10 indicating strong liking and 1 indicating little liking). The resulting dataset contains a liking value for each item in a trial, the participants' choice and RT, as well as the participant's gaze towards each item in a trial (describing the fraction of trial time that the participant spent looking at each item in the choice set).

To simulate individuals' response behaviour, we utilize the parameter estimates that

were obtained by Thomas et al. (2019) for the individuals in the three item choice dataset by Krajbich and Rangel (2011) (see Fig. A2). Importantly, we assume that ten individuals do not exhibit a gaze bias, meaning that their choices are independent of the time that they spend looking at each item. To this end, we set the γ value of ten randomly selected individuals to 1. We further assume that individuals' gaze is distributed randomly with respect to the values of the items in a choice set. An overview of the generating parameter estimates is given Fig. A1.

We first instantiate a GLAM model instance using `gb.GLAM()` and then use its `simulate_group` method. This method requires us to specify whether the individuals of the group are either simulated individually (and thereby independent of one another) or as part of a group with hierarchical parameter structure (where the individual model parameters are drawn from a group distribution, see below). For the former, the generating model parameters (indicated in the following as `gen_parameters`) are provided as a dictionary, containing a list of the individual participant values for each model parameter:

```
import glambox as gb
import numpy as np
glam = gb.GLAM()
no_bias_subjects = np.random.choice(a=gen_parameters['gamma'].size,
                                    size=10,
                                    replace=False)
gen_parameters['gamma'][no_bias_subjects] = 1
glam.simulate_group(kind='individual',
                    n_individuals=30,
                    n_trials=300,
                    n_items=3,
                    parameters=gen_parameters)
```

As this example is focused on the individual level, we can further create a summary table, describing individuals' response behaviour on three behavioural metrics, using the `aggregate_subject_level_data` function from the `analysis` module. The resulting table contains individuals' mean RT, their probability of choosing the item with the highest item value from a choice set and a behavioural measure of the strength of the association between individuals' gaze allocation and choice behaviour (indicating the mean increase in choice probability for an item that was fixated on longer than the others, after correcting for the influence of the item value on choice behaviour; for further details, see Thomas et al. 2019).

```
from glambox.analysis import aggregate_subject_level_data
subject_data_summary = aggregate_subject_level_data(data=glam.data,
                                                    n_items=3)
```

Exploring the behavioural data. In a first step of our analysis, we explore differences in individuals' response behaviour. To this end, we plot the distributions of individuals' scores on the three behavioural metrics, and their associations, using the `plot_behaviour_associations` function implemented in the `plots` module:

```
gb.plots.plot_behaviour_associations(data=data)
```

The resulting plot is displayed in Fig. 3 and shows that individuals' probability of choosing the best item, as well as the strength of their behavioural association of gaze and choice, are not associated with their mean RT (Fig. 3D-E). However, individuals' probability of choosing the best item increases with decreasing strength of the behavioural association of gaze and choice (Fig. 3F).

Likelihood-based model comparison. In a second step of our analysis, we want to test whether the response behaviour of each individual is better described by a decision model with or without gaze bias. To this end, we set up the two GLAM variants:

```
glam_bias = gb.GLAM(data=data)
glam_bias.make_model(kind='individual')

glam_nobias = gb.GLAM(data=data)
glam_nobias.make_model(kind='individual', gamma_val=1)
```

For the GLAM variant without gaze bias mechanism, we use the `gamma_val` argument and set it to a value of 1 (fixing γ to 1 for all subjects).

Subsequently, we fit both models to the data of each individual and compare their fit by means of the Widely Applicable Information Criterion (WAIC; Vehtari, Gelman, & Gabry, 2017):

```
glam_bias.fit(method='MCMC',
              tune=5000,
              draws=5000)

glam_nobias.fit(method='MCMC',
                tune=5000,
                draws=5000)
```

The fit method defaults to Metropolis-Hastings MCMC sampling (for methodological details, see Methods Section). The `draws` argument sets the number of samples to be drawn. This excludes the tuning (or burn-in) samples, which can be set with the `tune` argument. In addition, the `fit` method accepts the same keyword arguments as the PyMC3 sample function, which it wraps (see the PyMC3 documentation for additional details). The `chains` argument sets the number of MCMC traces (it defaults to four and should be set to at least two, in order to allow convergence diagnostics).

After convergence has been established for all parameter traces (for details on the suggested convergence criteria, see Methods), we perform a model comparison on the individual level, using the `compare` function of the PyMC3 library. For each individual, this function requires as input a dictionary, containing the individual's model and trace. The individually fitted models, as well as their traces, can be accessed through the `model` and `trace` attributes of our GLAM instances. Both are lists, with one entry per subject in the dataset. The `ic` argument specifies the information criterion to be used for the model comparison (`'WAIC'` or Leave-One-Out Cross Validation `'LOO'`). The `compare` function returns a table containing an estimate of the specified information criterion for each inputted model variant.

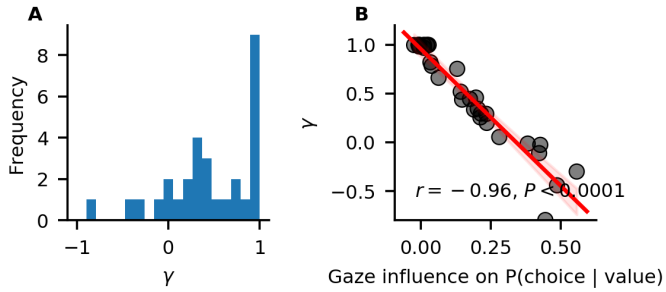

Figure 4. Individual differences in the strength of the association of gaze allocation and choice behaviour. A) Distribution of γ estimates resulting from the in-sample individual model fits. B) Association of γ estimates and individuals' values on the behavioural gaze bias measure. The red line indicates a linear regression fit, with surrounding 95% confidence bands. Pearson's r correlation with P-value is given.

```
from pymc3 import compare
n_subjects = 30
for s in range(n_subjects):
    waic_df = compare({glam_bias.model[s]:
                       glam_bias.trace[s],
                       glam_nobias.model[s]:
                       glam_nobias.trace[s]},
                       ic='WAIC')
```

With this comparison, we are able to identify those participants whose response behaviour matches the assumption of gaze-biased evidence accumulation. In particular, we find that we accurately recover whether an individual has a gaze bias or not for 29 out of 30 individuals.

Looking at the individual parameter estimates (defined as *maximum a posteriori* (MAP) of the posterior distributions), we find that the individually fitted γ values (Fig. 4A) cover a wide range between -0.8 and 1, indicating strong variability in the strength of individuals' gaze bias. We also find that γ estimates have a strong negative correlation with individuals' scores on the behavioural gaze bias measure (Fig. 4B).

Out-of-sample prediction. We have identified those participants whose response behaviour is better described by a GLAM variant with gaze-bias than one without. Yet, this analysis does not indicate whether the GLAM is a good model of individuals' response behaviour on an absolute level. To test this, we perform an out-of-sample prediction exercise.

We divide the data of each subject into even- and odd-numbered experiment trials and use the data of the even-numbered trials to fit both GLAM variants:

```
glam_bias.exchange_data(data_even)
glam_bias.fit(method='MCMC',
              tune=5000,
              draws=5000)
```

```
glam_nobias.exchange_data(data_even)
glam_nobias.fit(method='MCMC',
                tune=5000,
                draws=5000)
```

Subsequently, we evaluate the performance of both models in predicting individuals' response behaviour using the MAP estimates and item value and gaze data from the odd-numbered trials. To predict response behaviour for the odd-numbered trials, we use the `predict` method. We repeat every trial 50 times in the prediction (as specified through the `n_repeats` argument) to obtain a stable pattern of predictions:

```
glam_bias.exchange_data(data_odd)
glam_bias.predict(n_repeats=50)

glam_nobias.exchange_data(data_odd)
glam_nobias.predict(n_repeats=50)
```

Lastly, to determine the absolute fit of both model variants to the data, we plot the individually predicted against the individually observed data on all three behavioural metrics. To do this, we use the `plot_individual_fit` function of the `plots` module. This function takes as input the observed data, as well as a list of the predictions of all model variants that ought to be compared. The argument `prediction_labels` specifies the naming used for each model in the resulting figure. For each model variant, the function creates a row of panels, plotting the observed against the predicted data:

```
from glambox.plots import plot_individual_fit
plot_individual_fit(observed=data_odd,
                    predictions=[glam_bias.prediction,
                                 glam_nobias.prediction],
                    prediction_labels=['gaze-bias',
                                       'no gaze-bias'])
```

The resulting plot is displayed in Fig. 5. We find that both model variants perform well in capturing individuals' RTs and probability of choosing the best item (Fig. 5A, D, B, E). Importantly, only the GLAM variant with gaze bias is able to also recover the strength of the association between individuals' choice behaviour and gaze allocation (Fig. 5C).

Conclusion. GLAMbox provides an easy-to-use tool to test the presence (and variability) of gaze biases on the individual level. With GLAMbox, we can easily fit the GLAM to individual participant data, compare different model variants and predict individuals' response behaviour. It also provides a set of analysis functions to explore behavioural differences between individuals and to compare the fit of different model variants to observed response behaviour.

Example 2: Hierarchical Parameter Estimation in Cases with Limited Data

In some research settings, the total amount of data one can collect per individual is limited, conflicting with the large amounts of data required to obtain reliable and precise individual parameter estimates from diffusion models (Lerche, Voss, & Nagler, 2017; Voss,

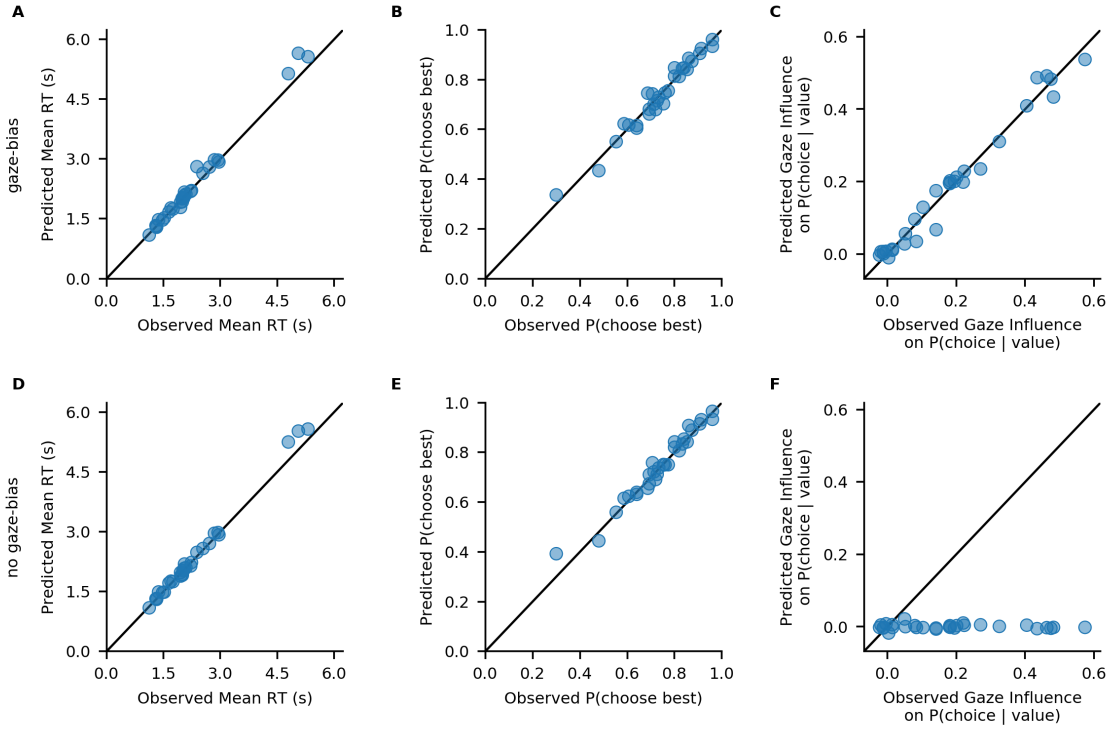

Figure 5. Comparison of individuals' observed response behaviour with the out-of-sample predictions of a GLAM variant with (A-C) and without gaze bias (D-F): Individuals' mean RT (A, D), probability of choosing the best item (B, E), and influence of gaze allocation on choice probability (C, F). Points indicate individual participant means.

Nagler, & Lerche, 2013). Hierarchical modeling can offer a solution to this problem. Here, each individual's parameter estimates are assumed to be drawn from a group level distribution. Thereby, during parameter estimation, individual parameter estimates are informed by the data of the entire group. This can greatly improve parameter estimation, especially in the face of limited amounts of data (Ratcliff & Childers, 2015; Wiecki et al., 2013). In this example, we will simulate a clinical application setting, in which different patient groups are to be compared on the strengths of their gaze biases, during a simple value-based choice task that includes eye tracking. It is reasonable to assume that the amount of data that can be collected in such a setting is limited on at least two accounts:

1. The number of patients available for the experiment might be low
2. The number of trials that can be performed by each participant might be low, for clinical reasons (e.g., patients feel exhausted more quickly, time to perform tests is limited, etc.)

Therefore, we simulate a dataset with a low number of individuals within each group (between 5 and 15 per group), and a low number of trials per participant (50 trials). We then estimate model parameters in a hierarchical fashion, and compare the group level gaze bias parameter estimates between groups.

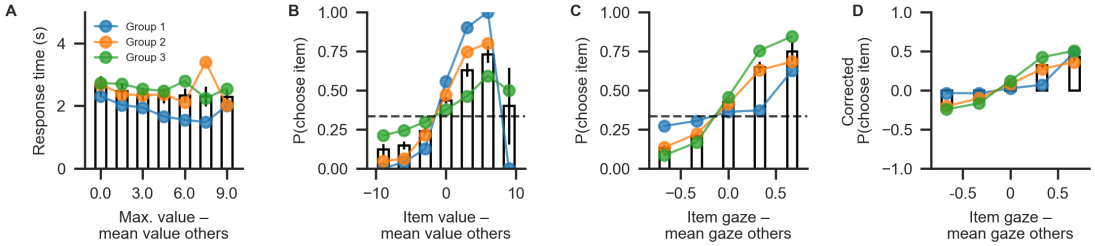

Figure 6. Aggregate view of the simulated data for Example 2. The data can be visualized quickly using the `plot_behaviour_aggregate` function from the `plots` module. It plots (A) mean RT, binned by trial difficulty (defined as the difference between the highest item value and the mean of all other item values), (B) the probability of choosing an item based on its relative value (defined as the difference between an item's value and the mean value of all other items), (C) the probability of choosing an item based on the relative gaze towards it (the difference between the gaze towards this item and the mean gaze towards all others) and (D) the probability of choosing an item, based on the relative gaze towards it, and corrected for the influence of item value. Bars correspond to data pooled across groups, coloured lines indicate individual groups.

Simulating data. We simulate data of three patient groups ($N_1 = 5$, $N_2 = 10$, $N_3 = 15$), with 50 trials per individual, in a simple three item value-based choice task, where participants are instructed to simply choose the item they like the best. These numbers are roughly based on a recent clinical study on the role of the prefrontal cortex in fixation-dependent value representations (Vaidya & Fellows, 2015). As before, we sample parameter sets for each individual from the estimates obtained from fitting the model to the data from (Krajbich & Rangel, 2011). This ensures that the simulated data used here closely resemble data that could be obtained experimentally. Importantly, we let the gaze bias parameter γ differ systematically between the groups, with means of $\gamma_1 = 0.7$ (weak gaze bias), $\gamma_2 = 0.1$ (moderate gaze bias) and $\gamma_3 = -0.5$ (strong gaze bias), respectively. All other parameters are sampled from the same distribution across groups (for an overview of the generating parameters, see Fig. A4). The groups primarily differ in the gaze bias parameter γ, whereas other parameters largely overlap (even though there is some non-systematic variance between individuals).

Behavioural differences between the three groups are plotted in Fig. 6, using the `plot_behaviour_aggregate` function from the `plots` module. Group-level summary tables can be created using the `aggregate_group_level_data` from the `analysis` module. Even though the groups only differ in the gaze bias parameter, they also exhibit differences in RT (Group 1 mean ± s.d. = 1.96 ± 0.33 s, Group 2 mean ± s.d. = 2.38 ± 1.4 s; Group 3 mean ± s.d. = 2.59 ± 1.26 ms; Fig. 6A) and choice accuracy (Group 1 mean ± s.d. = 0.88 ± 0.06, Group 2 mean ± s.d. = 0.71 ± 0.07, Group 3 mean ± s.d. = 0.50 ± 0.16; Fig. 6B). As is to be expected, we can also observe behavioural differences in gaze influence measure (Group 1 mean ± s.d. = 0.08 ± 0.07, Group 2 mean ± s.d. = 0.26 ± 0.11, Group 3 mean ± s.d. = 0.38 ± 0.11; Fig. 6 C-D, where the choices of Group 3 are driven by gaze more than those of the other groups.

Building the hierarchical model. When specifying the hierarchical model, we allow all model parameters to differ between the three groups. This way, we will subsequently be able to address the question whether individuals from different groups differ on one or more model parameters (including the gaze bias parameter γ, which we are mainly interested in here). As for the individual models, we first initialize the model object using the GLAM class and supply it with the behavioural data using the `data` argument. Here, we set the model kind to `'hierarchical'` (in contrast to `'individual'`). Further, we specify that each model parameter can vary between groups (referring to a `'group'` variable in the data):

```
hglam = gb.GLAM(data=data)
hglam.make_model(kind='hierarchical',
                 depends_on=dict(v='group',
                                 gamma='group',
                                 s='group',
                                 tau='group'))
```

Parameter estimation with MCMC. After the model is built, the next step is to perform statistical inference over its parameters. As we have done with the individual models, we can use MCMC to approximate the parameters' posterior distributions (see Methods for details):

```
hglam.fit(method='MCMC',
          draws=10000, tune=10000)
```

Evaluating parameter estimates, interpreting results. After sampling is finished, and the chains were checked for convergence, we can turn back to the research question: Do the groups differ with respect to their gaze biases? Questions about differences between group-level parameters can be addressed by inspecting their posterior distributions. For example, the probability that the mean $\gamma_{1,\mu}$ for Group 1 is larger than the mean $\gamma_{2,\mu}$ of Group 2 is given by the proportion of posterior samples in which this was the case.

GLAMbox includes a `plot_node_hierarchical` function that plots posterior distributions of group level parameters. Additionally, the user can specify a list of comparisons between groups or conditions. If comparisons are specified, the posterior distributions of their difference and corresponding relevant statistics are added to the figure:

```
from glambox.plots import plot_node_hierarchical
parameters = ['v', 'gamma', 's', 'tau']
comparisons = [('group1', 'group2'),
               ('group1', 'group3'),
               ('group2', 'group3')]
plot_node_hierarchical(model=hglam,
                       parameters=parameters,
                       comparisons=comparisons)
```

With the resulting plot (Fig. 7), the researcher can infer that the groups did not differ with respect to their mean velocity parameters $v_{i,\mu}$ (top row, pairwise comparisons), mean

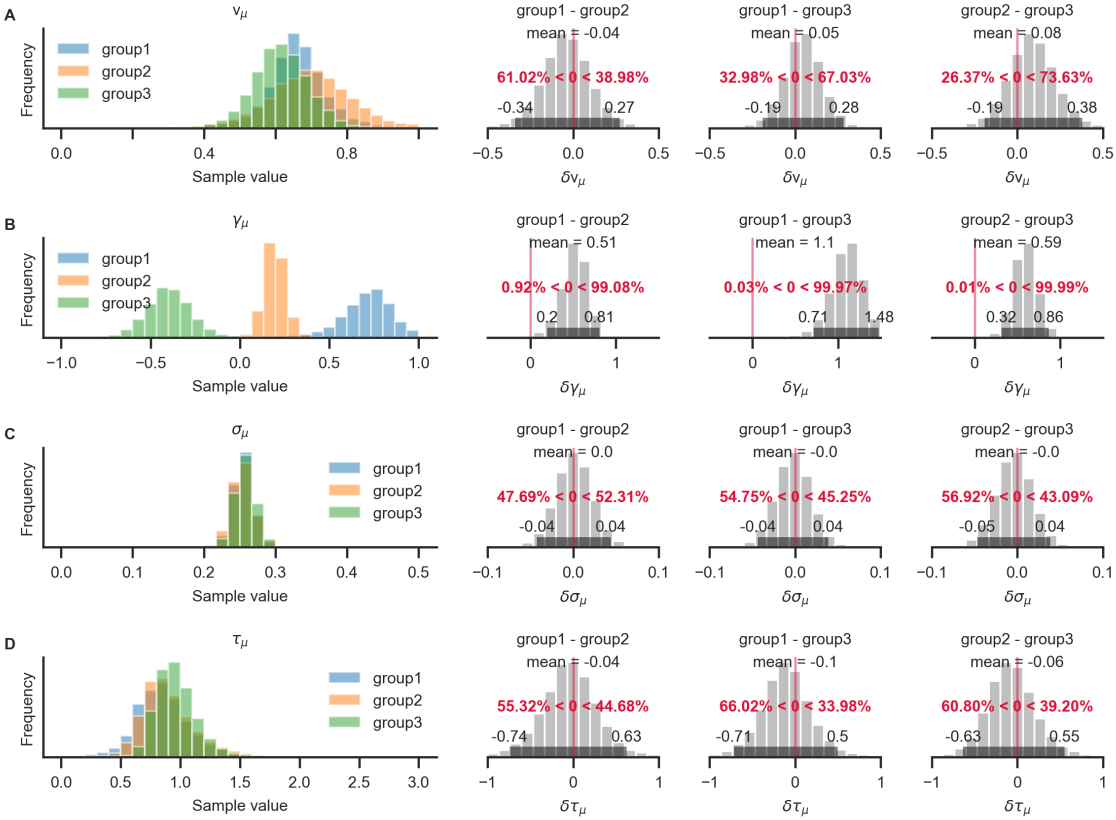

Figure 7. Pairwise comparison of posterior group-level parameter estimates between groups. The resulting figure has one row per model parameter. The leftmost column shows posterior distributions for each group. The next columns show posterior distributions of pairwise differences, as specified by the comparisons argument. These plots include the mean and 95% HPD values of the posterior distribution of the difference, as well as the proportion of samples below and above zero (in red). In this example, differences between all three groups on the γ parameter (row B) are present, as the posterior distributions of pairwise differences between groups clearly exclude zero. For example, the difference between groups 1 and 2 is estimated at 0.51, 95% HPD = [0.2, 0.81], with over 99% of the posterior mass above zero. No evidence for differences on any of the other model parameters is found (their posterior distributions largely overlap and the posterior distributions of pairwise differences between groups all clearly include zero).

accumulation noise $\sigma_{i,\mu}$ (third row), or scaling parameters $\tau_{i,\mu}$. The groups differ, however, in the strength of their mean gaze bias $\gamma_{i,\mu}$ (second row): All differences between the groups were statistically meaningful (as inferred by the fact that the corresponding 95% HPD did not contain zero; second row, columns 2-4).

Conclusion. When faced with limited data, GLAMbox allows users to easily build and estimate hierarchical GLAM variants, including conditional dependencies of model parameters. The Bayesian inference framework allows the researcher to answer relevant questions in a straightforward fashion. To this end, GLAMbox provides basic functions for

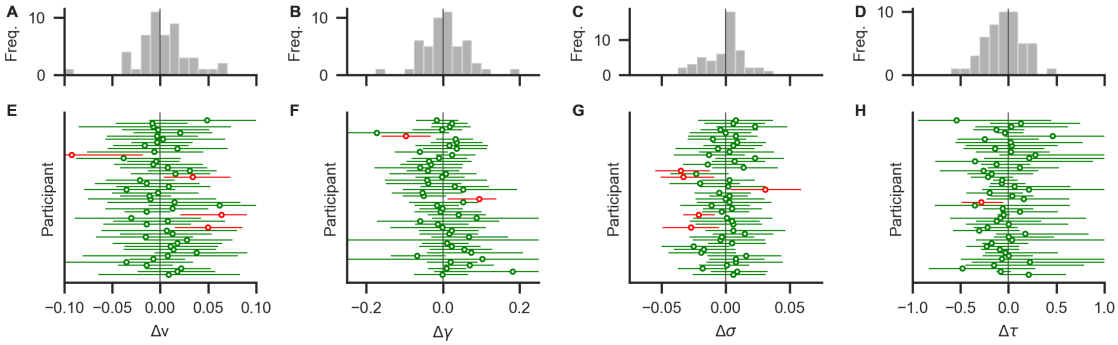

Figure 8. Results from a basic parameter recovery. The lower row (E-H) shows deviations between known generating parameter values and recovered MAP estimates (circles) and their 95% HPDs (horizontal error bars) for each participant. Green (red) colour indicates that the true value is within (outside) the 95% HPD. Most parameters were recovered with small deviations. Panels A-D show distributions of deviations across individuals. Distributions are mostly centered around zero, indicating no systematic under- or overestimation (bias) across individuals.

computation and visualization.

Example 3: Parameter Recovery

When performing model-based analyses of behaviour that include the interpretation of parameter estimates, or comparisons of parameter estimates between groups or conditions, the researcher should be confident that the model's parameters are actually identifiable. In particular, the researcher needs to be confident that the set of estimated parameters unambiguously describes the observed data better than any other set of parameters. A straightforward way of testing this is to perform a parameter recovery: The general intuition of a parameter recovery analysis is to first generate a synthetic dataset from a model using a set of known parameters, and then fitting the model to the synthetic data. Finally, the estimated parameters can be compared to the known generating parameters. If they match to a satisfying degree, the parameters were recovered successfully. Previous analyses have already indicated that the GLAM's parameters can be recovered to a satisfying degree (Thomas et al., 2019). Yet, the ability to identify a given set of parameters always depends on the specific features of a given dataset. The most obvious feature of a dataset that influences recoverability of model parameters is the number of data points included. Usually this quantity refers to the number of trials that participants performed. For hierarchical models, the precision of group-level estimates also depends on the number of individuals per group. Additional features that vary between datasets and that could influence parameter estimation are the observed distribution of gaze, the distribution of item values or the number of items in each trial. For this reason, it is recommended to test whether the estimated parameters of a model can be recovered in the context of a specific dataset.

To demonstrate the procedure of a basic parameter recovery analysis using GLAMbox, suppose we have collected and loaded a data set called `data`. In the first step, we perform

parameter estimation as in the previous examples:

```
glam = gb.GLAM(data=data)
glam.make_model(kind='individual')
glam.fit(method='MCMC', draws=5000, tune=5000)
```

The next step is to create a synthetic, model-generated dataset using the model parameters estimated from the empirical data, together with the empirically observed stimulus and gaze data using the `predict` method. Setting `n_repeats` to 1 results in a dataset of the same size as the observed one:

```
glam.predict(n_repeats=1)
synthetic = glam.prediction
```

The synthetic dataset should resemble the empirically observed data closely. If there are major discrepancies between the synthetic and observed data, this indicates that GLAM might not be a good candidate model for the data at hand. Next, we create a new model instance, attach the synthetic data, build a model and re-estimate its parameters:

```
glam_rec = gb.GLAM(data=synthetic)
glam_rec.make_model(kind='individual')
glam_rec.fit(method='MCMC',
             draws=5000,
             tune=5000)
```

Finally, the recovered and generating parameters can be compared. If the recovered parameters do not match the generating parameters, the parameters cannot be identified given this specific dataset. In this case, parameter estimates should not be interpreted.

If, on the other hand, generating and recovered parameters do align, the parameters have been recovered successfully. This indicates that the model's parameters can be identified unambiguously given the general characteristics of the dataset and thereby increases confidence that the parameters obtained from the empirical data are valid and can be interpreted.

Here, all parameters could be recovered as illustrated in Fig. 8. For most individuals, the MAP estimates and their 95% HPDs are close to the known generating parameters. Across individuals, no systematic biases in the estimation can be identified.

Conclusion. In this example, we demonstrated how to perform a basic parameter recovery for a given dataset. When successful, this increases confidence that the parameters can be identified with the given dataset.

Discussion

Researchers have recently started to systematically investigate the role of visual gaze in the decision making process. By now, it is established that eye movements do not merely serve to sample information that is then processed independently to produce a choice, but that they are actively involved in the construction of preferences (Ashby et al., 2016; Cavanagh et al., 2014; Folke et al., 2017; Konovalov & Krajbich, 2016; Krajbich et al., 2010;

Krajbich & Rangel, 2011; Orquin & Mueller Loose, 2013; Shimojo et al., 2003; Tavares et al., 2017; Thomas et al., 2019). The dominant theoretical perspective is that evidence accumulation in favor of each option is modulated by gaze allocation, so that accumulation for non-fixated options is attenuated. This mechanism is formally specified in various models of gaze-dependent decision making, such as the attentional Drift Diffusion Model (aDDM; Krajbich et al., 2010; Krajbich & Rangel, 2011) and the conceptually related Gaze-weighted Linear Accumulator Model (GLAM; Thomas et al., 2019). In contrast to analyses based on behavioural and eye tracking data alone, these models can act as analytical tools that enable researchers to address questions regarding specific mechanisms in the decision process, like the gaze bias. They further formally establish mechanistic associations between choice, RT and eye tracking data and enable prediction of these data. Even though the advantages of applying these models are apparent, their use is limited by their complexity and the high cost of implementing, validating and optimizing them. Further, there are only few off-the-shelf solutions researchers can turn to, if they want to perform model-based analyses of gaze-dependent choice data, particularly for choice settings involving more than two alternatives. With GLAMbox, we present a Python-based toolbox, built on top of PyMC3, that allows researchers to perform model-based analyses of gaze-bias effects in decision making easily. We have provided step-by-step instructions and code to perform essential modeling analyses using the GLAM. These entail application of the GLAM to individual and group-level data, specification of parameter dependencies for both within- and between-subject designs, (hierarchical) Bayesian parameter estimation, comparisons between multiple model variants, out-of-sample prediction of choice and RT data, data visualization, Bayesian comparison of posterior parameter estimates between conditions, and parameter recovery. We hope that GLAMbox will make studying the association between gaze allocation and choice behaviour more accessible. We also hope that the resulting findings will ultimately help us better understand this association, its inter-individual variability and link to brain activity.

References

Armel, K. C., Beaumel, A., & Rangel, A. (2008, June). Biasing simple choices by manipulating relative visual attention. *Judgment and Decision Making*, *3*, 396–403. Retrieved 2019-07-15, from http://resolver.caltech.edu/CaltechAUTHORS:ARMjdm08

Ashby, N. J. S., Jekel, M., Dickert, S., & Glöckner, A. (2016). Finding the right fit: A comparison of process assumptions underlying popular drift-diffusion models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(12), 1982–1993. doi: 10.1037/xlm0000279

Cavanagh, J. F., Wiecki, T. V., Kochar, A., & Frank, M. J. (2014). Eye tracking and pupillometry are indicators of dissociable latent decision processes. *Journal of Experimental Psychology: General*, *143*(4), 1476–1488. doi: 10.1037/a0035813

Fiedler, S., & Glöckner, A. (2012). The Dynamics of Decision Making in Risky Choice: An Eye-Tracking Analysis. *Frontiers in Psychology*, *3*. Retrieved 2018-11-20, from https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00335/full doi: 10.3389/fpsyg.2012.00335

Fisher, G. (2017, November). An attentional drift diffusion model over binary-attribute choice. *Cognition*, *168*, 34–45. Retrieved 2018-11-30, from http://www.sciencedirect.com/science/article/pii/S0010027717301695 doi: 10.1016/j.cognition.2017.06.007

Folke, T., Jacobsen, C., Fleming, S. M., & De Martino, B. (2017, January). Explicit representation of confidence informs future value-based decisions. *Nature Human Behaviour*, *1*(1), 0002.

Retrieved 2019-01-16, from https://www.nature.com/articles/s41562-016-0002 doi: 10.1038/s41562-016-0002

Gamerman, D., & Lopes, H. F. (2006). *Markov Chain Monte Carlo : Stochastic Simulation for Bayesian Inference, Second Edition*. Chapman and Hall/CRC. Retrieved 2019-07-15, from https://www.taylorfrancis.com/books/9780429183348 doi: 10.1201/9781482296426

Gelman, A., & Shirley, K. (2011). Inference from simulations and monitoring convergence. *Handbook of markov chain monte carlo*, *6*, 163–174.

Glöckner, A., & Herbold, A.-K. (2011). An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. *Journal of Behavioral Decision Making*, *24*(1), 71–98. Retrieved 2019-01-24, from https://onlinelibrary.wiley.com/doi/abs/10.1002/bdm.684 doi: 10.1002/bdm.684

Konovalov, A., & Krajbich, I. (2016, August). Gaze data reveal distinct choice processes underlying model-based and model-free reinforcement learning. *Nature Communications*, *7*, 12438. Retrieved 2019-01-24, from https://www.nature.com/articles/ncomms12438 doi: 10.1038/ncomms12438

Krajbich, I., Armel, C., & Rangel, A. (2010, October). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298. Retrieved 2019-01-24, from https://www.nature.com/articles/nn.2635 doi: 10.1038/nn.2635

Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The Attentional Drift-Diffusion Model Extends to Simple Purchasing Decisions. *Frontiers in Psychology*, *3*. Retrieved 2018-11-20, from https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00193/full doi: 10.3389/fpsyg.2012.00193

Krajbich, I., & Rangel, A. (2011, August). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, *108*(33), 13852–13857. Retrieved 2019-01-24, from https://www.pnas.org/content/108/33/13852 doi: 10.1073/pnas.1101328108

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.

Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A., & Blei, D. M. (2017). Automatic differentiation variational inference. *The Journal of Machine Learning Research*, *18*(1), 430–474.

Lerche, V., Voss, A., & Nagler, M. (2017, April). How many trials are required for parameter estimation in diffusion modeling? A comparison of different optimization criteria. *Behavior Research Methods*, *49*(2), 513–537. Retrieved 2019-03-07, from https://doi.org/10.3758/s13428-016-0740-2 doi: 10.3758/s13428-016-0740-2

Lopez-Persem, A., Domenech, P., & Pessiglione, M. (2016, November). How prior preferences determine decision-making frames and biases in the human brain. *eLife*, *5*, e20317. Retrieved 2019-01-24, from https://doi.org/10.7554/eLife.20317 doi: 10.7554/eLife.20317

McKinney, W. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th python in science conference* (Vol. 445, pp. 51–56).

Orquin, J. L., & Mueller Loose, S. (2013, September). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, *144*(1), 190–206. Retrieved 2018-12-18, from http://www.sciencedirect.com/science/article/pii/S0001691813001364 doi: 10.1016/j.actpsy.2013.06.003

Pärnamets, P., Johansson, P., Hall, L., Balkenius, C., Spivey, M. J., & Richardson, D. C. (2015, March). Biasing moral decisions by exploiting the dynamics of eye gaze. *Proceedings of the National Academy of Sciences*, *112*(13), 4170–4175. Retrieved 2019-01-24, from https://www.pnas.org/content/112/13/4170 doi: 10.1073/pnas.1415250112

Ratcliff, R., & Childers, R. (2015). Individual Differences and Fitting Methods for the Two-Choice Diffusion Model of Decision Making. *Decision (Washington, D.C.)*, *2015*. Retrieved 2019-03-07, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4517692/

Ratcliff, R., & Tuerlinckx, F. (2002, September). Estimating parameters of the diffusion model: Ap-

proaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, *9*(3), 438–481. Retrieved 2019-01-24, from https://doi.org/10.3758/BF03196302 doi: 10.3758/BF03196302

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016, April). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, *2*, e55. Retrieved 2019-01-24, from https://peerj.com/articles/cs-55 doi: 10.7717/peerj-cs.55

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003, December). Gaze bias both reflects and influences preference. *Nature Neuroscience*, *6*(12), 1317–1322. Retrieved 2019-01-24, from https://www.nature.com/articles/nn1150 doi: 10.1038/nn1150

Smith, S. M., & Krajbich, I. (2018, December). Attention and choice across domains. *Journal of Experimental Psychology: General*, *147*(12), 1810–1826. Retrieved 2018-12-18, from http://search.ebscohost.com/login.aspx?direct=true&db=pdh&AN=2018-47076-001&site=ehost-live doi: 10.1037/xge0000482

Stewart, N., Gächter, S., Noguchi, T., & Mullett, T. L. (2016). Eye Movements in Strategic Choice. *Journal of Behavioral Decision Making*, *29*(2-3), 137–156. Retrieved 2019-01-24, from https://onlinelibrary.wiley.com/doi/abs/10.1002/bdm.1901 doi: 10.1002/bdm.1901

Stewart, N., Hermens, F., & Matthews, W. J. (2016, April). Eye Movements in Risky Choice. *Journal of Behavioral Decision Making*, *29*(2-3), 116–136. Retrieved 2019-01-24, from https://onlinelibrary.wiley.com/doi/full/10.1002/bdm.1854 doi: 10.1002/bdm.1854

Tavares, G., Perona, P., & Rangel, A. (2017). The Attentional Drift Diffusion Model of Simple Perceptual Decision-Making. *Frontiers in Neuroscience*, *11*. Retrieved 2018-11-20, from https://www.frontiersin.org/articles/10.3389/fnins.2017.00468/full doi: 10.3389/fnins.2017.00468

Thomas, A. W., Molter, F., Krajbich, I., Heekeren, H. R., & Mohr, P. N. C. (2019, April). Gaze bias differences capture individual choice behaviour. *Nature Human Behaviour*, 1. Retrieved 2019-04-16, from https://www.nature.com/articles/s41562-019-0584-8 doi: 10.1038/s41562-019-0584-8

Trueblood, J. S., Brown, S. D., Heathcote, A., & Busemeyer, J. R. (2013, June). Not Just for Consumers: Context Effects Are Fundamental to Decision Making. *Psychological Science*, *24*(6), 901–908. Retrieved 2019-04-12, from https://doi.org/10.1177/0956797612464241 doi: 10.1177/0956797612464241

Vaidya, A. R., & Fellows, L. K. (2015, December). Testing necessary regional frontal contributions to value assessment and fixation-based updating. *Nature Communications*, *6*, 10120. Retrieved 2019-01-24, from https://www.nature.com/articles/ncomms10120 doi: 10.1038/ncomms10120

Vehtari, A., Gelman, A., & Gabry, J. (2017, September). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. Retrieved 2019-01-24, from https://doi.org/10.1007/s11222-016-9696-4 doi: 10.1007/s11222-016-9696-4

Voss, A., Nagler, M., & Lerche, V. (2013, January). Diffusion Models in Experimental Psychology. *Experimental Psychology*, *60*(6), 385–402. Retrieved 2019-03-07, from https://econtent.hogrefe.com/doi/full/10.1027/1618-3169/a000218 doi: 10.1027/1618-3169/a000218

Voss, A., & Voss, J. (2007, November). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods*, *39*(4), 767–775. Retrieved 2019-07-15, from https://doi.org/10.3758/BF03192967 doi: 10.3758/BF03192967

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, *7*. Retrieved 2019-01-24, from https://www.frontiersin.org/articles/10.3389/fninf.2013.00014/full doi: 10.3389/fninf.2013.00014

Acknowledgments

The Junior Professorship of P.N.C.M. as well as the associated Dahlem International Network Junior Research Group Neuroeconomics is supported by Freie Universität Berlin within the Excellence Initiative of the German Research Foundation (DFG). Further support for P.N.C.M. is provided by the WZB Berlin Social Science Center. F.M. is supported by the International Max Planck Research School on the Life Course (LIFE). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

F.M. and A.W.T. contributed equally to the manuscript and share first authorship. F.M. and A.W.T. conceived of the GLAM, implemented all analyses and visualizations. F.M. and A.W.T. also co-wrote all software. F.M. and A.W.T. wrote the original draft of the manuscript, and H.R.H. and P.N.C.M. reviewed and edited the manuscript. Funding for this work was acquired by P.N.C.M. The work was supervised by H.R.H. and P.N.C.M.

Appendix

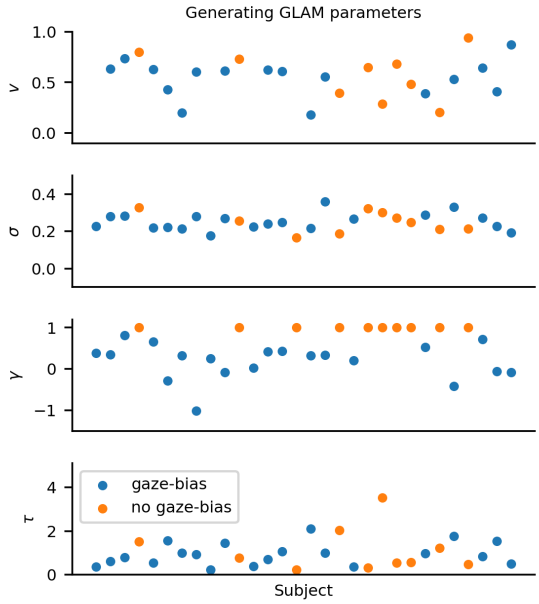

Figure A1. Distribution of individual generating GLAM parameters of Example 1. Colours indicate whether a subject was simulated with or without gaze bias.

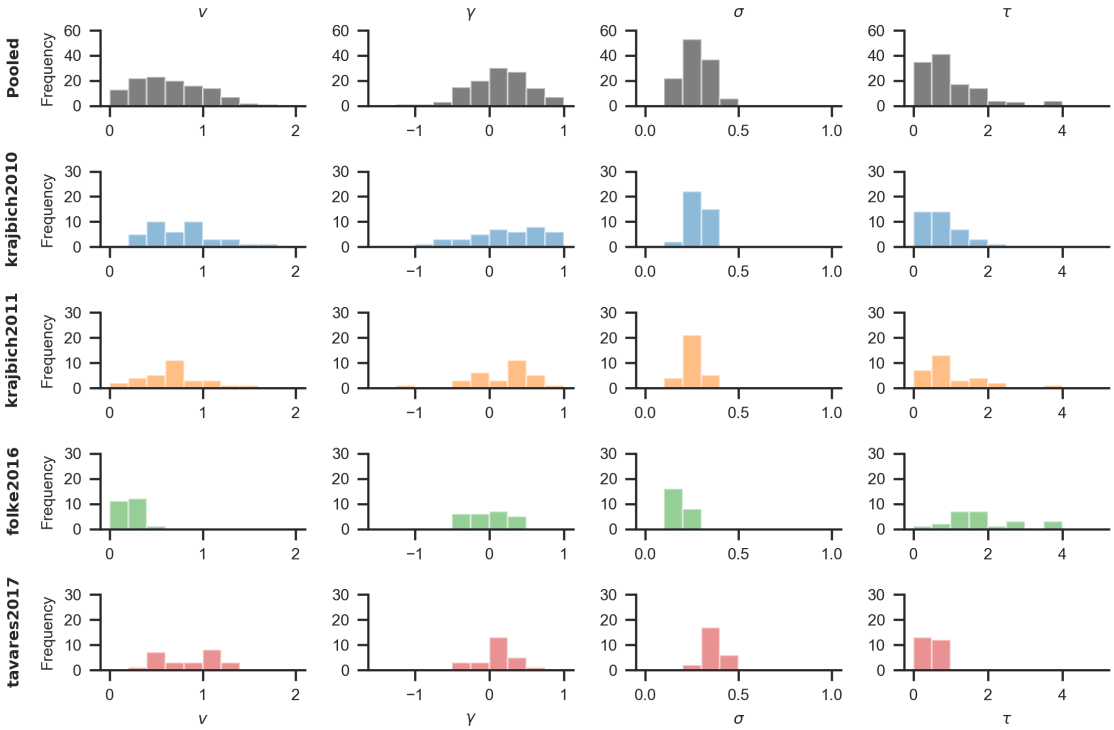

Figure A2. Distribution of individual parameter estimates from four datasets analysed in Thomas et al. (2019). The top row contains distributions of parameter estimates across datasets. Subsequent rows show distributions per dataset: Krajbich et al. (2010; blue), Krajbich & Rangel (2011; orange), Experiment 2 from Folke et al. (2017; green) and Experiment 1 from Tavares et al. (2017; red).

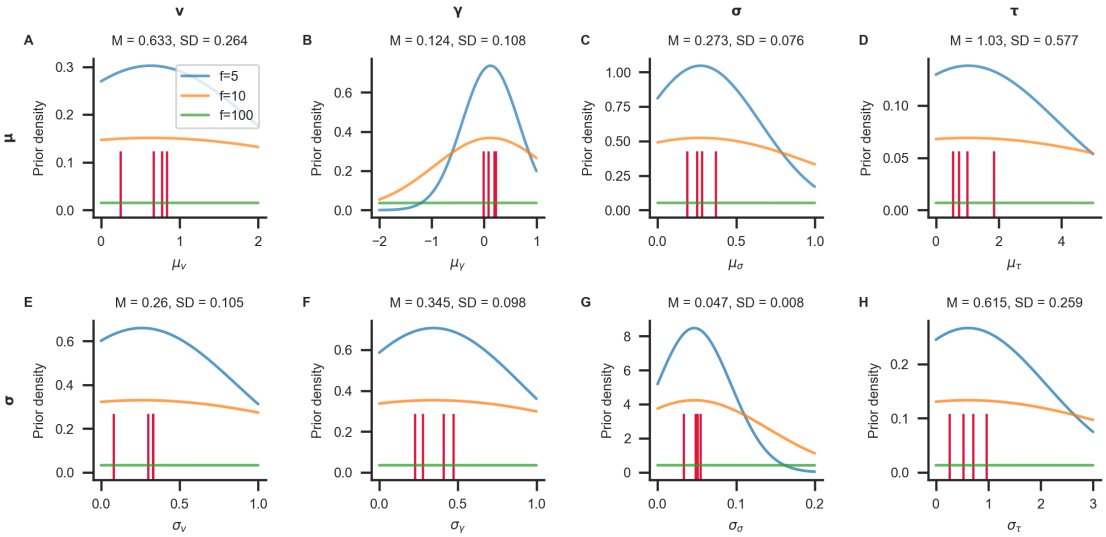

Figure A3. Illustration of hyperpriors. Different hyperpriors based on group-averaged parameter values were obtained from fitting the model to four different datasets (Folke et al., 2017; Krajbich et al., 2010; Krajbich & Rangel, 2011; Tavares et al., 2017; see Table A1 and Fig. A2). Panels show prior distributions on group level mean (upper row) and standard deviation (lower row) for each model parameter (columns; from left to right: v, γ, σ, τ). Observed group level estimates from the four datasets are indicated as red ticks in each panel. Blue, orange and green lines represent prior distributions with increasing levels of vagueness f. They are constructed as normal distributions with mean equal to the mean of the observed group level parameters across datasets (M), and standard deviation equal to f times the observed standard deviation across datasets (SD). Higher values of f correspond to wider, less informative priors. Prior distributions are further bounded between sensible limits. The user can specify the factor f during specification of hierarchical models. By default, hyperpriors with $f = 10$ (orange lines) are used.

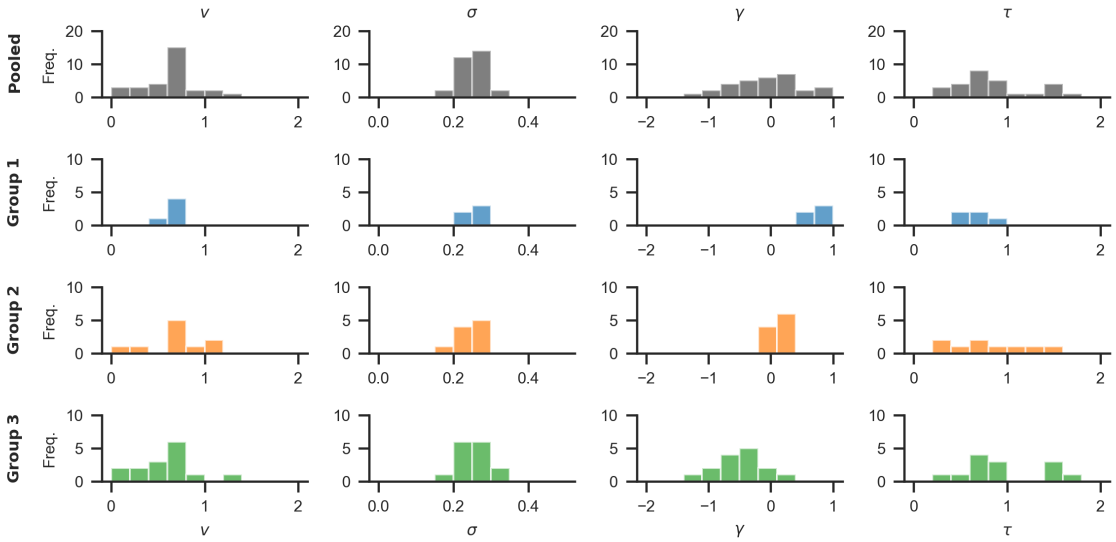

Figure A4. Distributions of data-generating parameters for the three groups in Example 2. The top row shows distributions pooled across groups. The bottom three rows show distributions per group. Note that the groups do not differ systematically with respect to the velocity parameter v, the noise parameter σ, or the scaling parameter τ (first, second and last column; even though there is some variability between individuals). The groups differ, however, on the gaze bias parameter γ (third column): Group 1 only has a weak gaze bias (large γ), group 2 has a medium strong gaze bias (smaller γ), and group 3 has a very strong gaze bias (even smaller, negative γ).

Dataset	N	Parameter	Mean	SD	min	25%	50%	75%	max
Folke 2016	24	v	0.25	0.08	0.17	0.19	0.22	0.3	0.51
		γ	-0.01	0.28	-0.45	-0.22	-0.04	0.2	0.43
		σ	0.19	0.03	0.12	0.17	0.19	0.21	0.28
		τ	1.85	0.97	0.43	1.19	1.62	2.52	3.92
Krajbich 2010	39	v	0.77	0.33	0.24	0.53	0.77	0.93	1.69
		γ	0.23	0.47	-0.78	-0.07	0.33	0.6	0.85
		σ	0.28	0.05	0.19	0.25	0.29	0.32	0.38
		τ	0.74	0.52	0.1	0.37	0.57	1.09	2.27
Krajbich 2011	30	v	0.67	0.33	0.18	0.44	0.63	0.78	1.58
		γ	0.19	0.41	-1.02	-0.07	0.33	0.45	0.81
		σ	0.25	0.05	0.17	0.22	0.25	0.28	0.36
		τ	1.0	0.71	0.24	0.52	0.81	1.39	3.53
Tavares 2017	25	v	0.84	0.3	0.38	0.54	0.9	1.02	1.4
		γ	0.09	0.23	-0.36	0.01	0.08	0.22	0.54
		σ	0.37	0.05	0.23	0.35	0.37	0.39	0.46
		τ	0.54	0.26	0.12	0.34	0.49	0.74	0.97

Table A1

Description of individual parameter estimates from four datasets analysed in Thomas et al. (2019). The datasets are originally from Folke et al., 2017 (Experiment 2); Krajbich et al., 2010; Krajbich & Rangel, 2011 and Tavares et al., 2017 (Experiment 1).