One direction? A tutorial for circular data in cognitive psychology.

Jolien Cremers [1]*, Irene Klugkist* [1,2]
[1]*Department of Methodology and Statistics, Utrecht University*
[2]*Research methodology, Measurement and Data Analysis, Behavioural Sciences, University of Twente*

1 Introduction

Circular data arises in almost all fields of research, from ecology where data on the direction movement of animals is investigated [1] to the medical sciences where protein structure [2] or neuronal activity [3] is investigated using periodic and thus circular measurements. The most direct examples of circular data within the social sciences arise in cognitive and experimental psychology. For example, in experiments on cognitive maps the human sense of direction is investigated through asking participants in a study to point north [4] or to walk to a certain object [5]. The closer the participants' pointing or walking direction was to the actual north or target object, the better their sense of direction. Other examples include the visual perception of space [6], visual working memory [7] and sensorimotor synchronization in music making [8].

However, despite the fact that circular data is being collected in different areas of cognitive and experimental psychology, the knowledge of this type of data is not well-spread. Circular data is fundamentally different from linear data due to its periodic nature. On the circle, measurements at $0°$ and $360°$ represent the same direction whereas on a linear scale they would be located at opposite ends of a scale. For this reason circular data require specific analysis methods. Some less technical textbooks on analysis methods for circular data have been written [9–11]. However these works are not part of the 'standard' texts on statistical analysis in psychology or the social sciences in general nor are they very well known amongst social scientific researchers.

Therefore, this paper aims at giving a tutorial in working with and analysing circular data to researchers in cognitive psychology and the social sciences in general. The main goal of this tutorial is to explain how to inspect and analyse your data when the outcome variable is circular. We will discuss data inspection, model fit, estimation and hypothesis testing in general linear models (GLM) and mixed-effects models. In this tutorial we decide to mainly focus on one particular approach to the analysis of circular data, the embedding approach. We do so for the flexibility of this approach and the resulting variety in types of models that have already been outlined in the literature on circular data for this approach. Note that for an optimal understanding of the paper, the reader should ideally have some knowledge on `R` and on multiple regression and mixed-effects models in the linear setting. The reader does not need to be familiar with circular data.

The structure of the tutorial is such that the reader is guided by two examples throughout the paper. One is an example for an ANOVA model and the other for a mixed-effects model. First however, we give a short introduction to circular data in general in Section 2. Next, in Section 3, we introduce the ANOVA example after which descriptive methods for circular data are explained through a section on data inspection for this example. After that we will continue with an analysis of the example datasets. First we analyse the ANOVA dataset using a method for circular GLM and give interpretation guidelines for this model in Section 4. Next, in Section 5 we will introduce
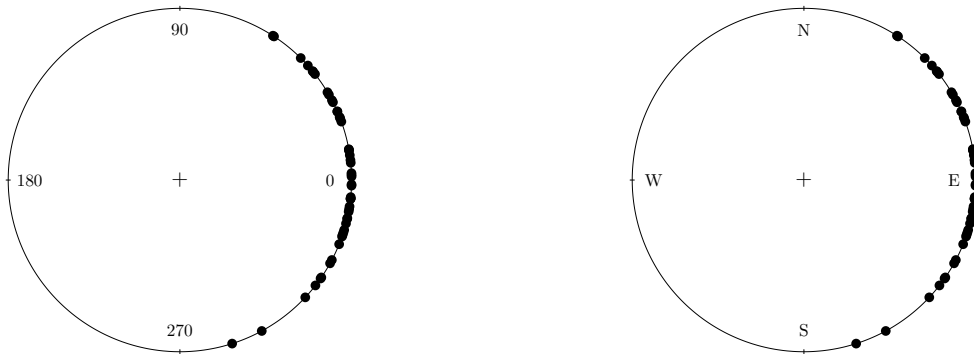
Figure 1: Data from participants in an experiment that were instructed to move East. The plot on the left shows the data on a 0° - 360° scale. The plot on the right shows the data on the compass.

and analyse the mixed-effects example data. Again, we analyse this data and include guidelines for interpretation. The analyses of both datasets, the ANOVA and mixed-effects dataset, are performed using the package `bpnreg`. For both models we use, the GLM and the mixed-effects model for a circular outcome we write a short technical section (sections 4.1 and 5.3.1) in which the mathematical details of the respective models are given. Lastly, a summary of the paper and additional references to literature on other models for circular data are given in Section 6.

2 Circular data

In the introduction we have briefly mentioned that circular data is data that has a periodic nature. The most intuitive form of circular data comes in the form of directions on a compass. For example, a participant in an experiment could be instructed to move or point to a certain target. We can then measure the direction, North, South, East or West on a scale from 0 to 360 degrees. A plot with simulated data containing such measurements for several participants is shown in Figure 1. In this plot we can easily see the periodicity of the data, 0° represents the same datapoint as 360°. Furthermore, we can see what happens if we would treat this data in the 'usual' linear way. Participants that moved North-East have a score of 45° and participants that moves South-East have a score of 315°. On the circle we can see that this is only 90° apart, while on a linear scale it is much further apart at 315° - 45° = 270°. More importantly however, there is a difference between the circular and linear means for this data. In Figure 1 we see that the circular mean direction is 0°. The linear mean however is 180° and is opposite to the actual mean direction of the data. Clearly, a linear treatment of the data in Figure 1 can lead to incorrect conclusions.

Clock times are another type of circular data. We might for instance be interested at what time of day a certain event takes place, e.g. the time of day at which positive affect is highest. Figure 2 shows
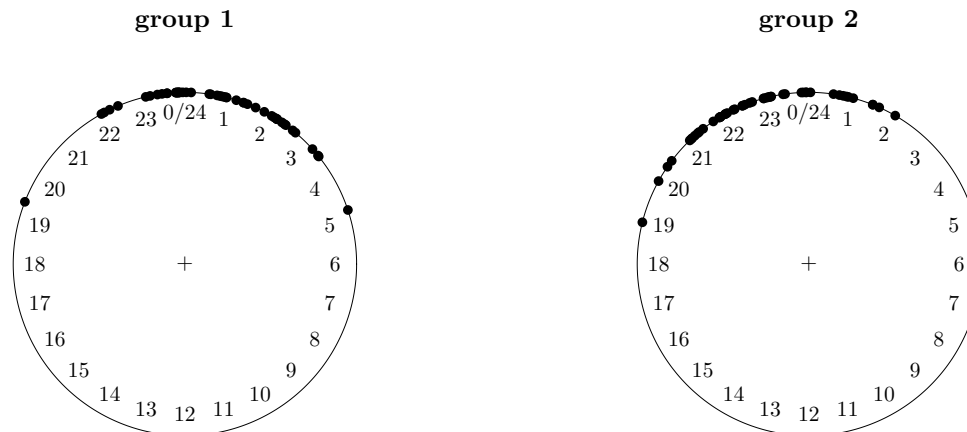
Figure 2: Data for the hour at which positive affect is highest for two groups of psychiatrical patients who are being treated for depression at different clinics.

simulated data for the time of day at which positive affect is highest for two groups of participants, e.g. two groups of psychiatrical patients who are being treated for depression at different clinics. From the plots we clearly see that the peak of positive affect for the two groups is at roughly the same time of day, one slightly before 12 pm and one slighly after. However, if we were to analyze this data using standard statistics for linear data and we would compare the means of the two groups, 11 pm and 1 am we would reach a completely different conclusion. The two means are namely at the two opposite ends of a linear scale from 00.00 am to 12.00 pm, and we would conclude that the time of day at which positive affect is highest is different for the two groups.

The two examples of circular data that we have just given illustrate why it is important to treat circular data differently from linear data. This goes both for describing your data, e.g. computing circular means as well as analyzing them, e.g. testing whether the circular means of two groups differ. In the next section we will introduce an example dataset on which we will show several ways to inspect and compute descriptive measures for circular data.

3 Inspecting your data

In the previous section we have seen that the computation of a circular mean differs from that of a linear mean. Also other methods for data inspection, the computation of other descriptive statistics and plotting methods, are different for circular data. Because data inspection shoud be done before performing inference of any kind we will outline a basic way to inspect circular data using the R [12] packages `bpnreg` [13] and `circular` [14]. We will discuss plots and several descriptive measures for circular data using an example dataset, the motor resonance data.

3.1 The motor resonance data

In this section we introduce data from an article by [15] on human motor resonance. From now on we will call this data the motor resonance data. Motor resonance is a response in the brain in the primary motor cortex and spinal circuits that is caused by observation of others' actions. In their research [15] conduct an experiment in which 'observers' are asked to either look at the movement of a hand of a 'mover' or at another object in order to evaluate the role of attention in motor resonant response. The experiment has three conditions: the 'explicit observation' condition (n = 14), where observers are explicitly instructed to observe the hand, the 'semi-implicit observation' condition (n = 14) where the observers have to perform a task that requires the implicit observation of the hand of the mover and the 'implicit observation' condition (n = 14), where observers have to perform a task that is independent of the observation of the hand of the mover. The idea of motor resonance is that the 'observer' starts moving his or her hand in the same manner as the 'mover' because he or she is implicitly or explicitly observing the hand of the 'mover'. This is the resonant response. This resonant response is hypothesized to be strongest and more synchronized with the hand movement of the mover in the explicit condition and smallest in the implicit condition. In each condition the hand movements of the observers were measured and the phase difference between movement of the observers' hand and the hand of the mover was calculated. This phase difference is a measurement of the strength of the resonant response and a circular variable. It can thus be described and analyzed using circular statistics. In addition to the phase difference the average amplitute of the hand movement of the observer was computed. Note that in the original article there was also a baseline condition (n=14) without a mover. Instead, observers had to look at an inanimate object that moves in an identical manner to the hand of the mover in other conditions. The baseline condition is however not included in the example data since no resonant response was observerd in the observers' hand according to the original research.

The motor resonance data can be found in the package `bpnreg` as the dataframe `Motor`. `Motor` is a dataframe with 42 rows and 7 variables. The variable `Cond` indicates the condition (explicit, semi-implicit and implicit) a participant was placed in, the `AvAmp` variable contains the average amplitude, and the `PhaseDiff` and `Phaserad` variables contain the measured phase difference between 'observer' and 'mover' in degrees and radians respectively. Note that circular data can be represented both in degrees on a scale from 0° to 360° and in radians on a scale from 0 to 2π (1 degree = 1 * $\pi/180$° radians).

3.2 Plots for circular data

The main question of interest for the motor resonance data is whether the phase difference between the three experimental conditions differs. To be more precise whether there is a smaller phase difference in the explicit condition than in the other two. Differences that are observed in the phase difference are interpreted as differences in the strength of the resonant response [15]. A smaller phase difference indicates a stronger and more synchronized resonant response. A first step to investigating the question of interest is plotting the phase differences of the three conditions. We can do so using the package `circular`.

Figure 3 shows the plots of the phase differences of each condition. We see that the phase differences in the explicit condition are much less spread out on the circle than the phase differences in the other two conditions. Also the average phase differences seem to differ between the conditions. In the next section we show measures for the mean and variance of a sample of circular data.
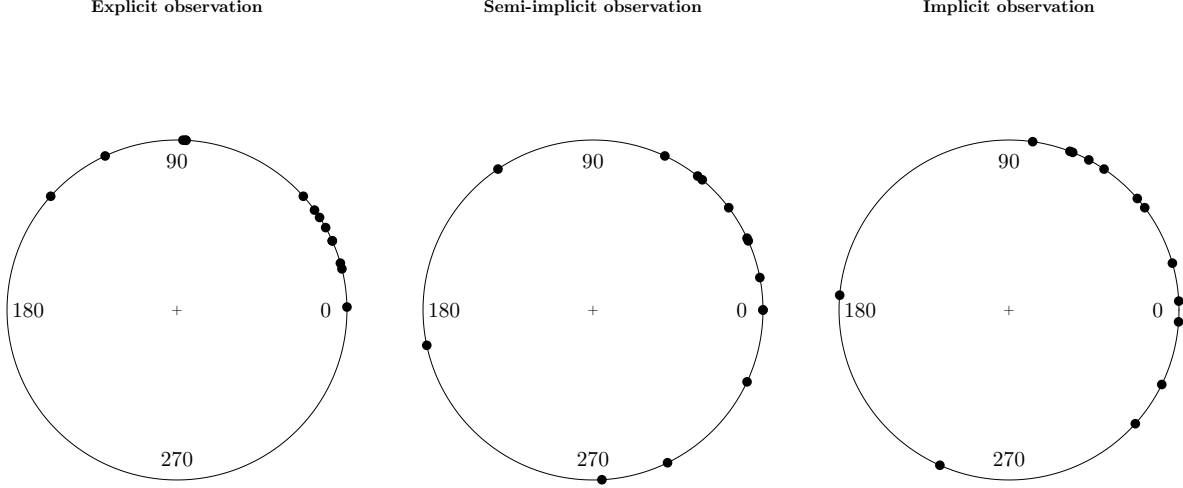
Figure 3: Plots of the phase differences for each condition of the motor resonance data.

Table 1: Descriptives for the motor resonance data with mean direction ($\bar{\theta}$), mean resultant length ($\bar{R}$), circular variance (V_m) and circular standard deviation (v) of the phase difference for each condition.

Phase difference	$\bar{\theta}$	$\bar{R}$	V_m	v
explicit	49.55°	0.77	0.23	41.39°
semi.implicit	18.45°	0.54	0.46	63.82°
implicit	31.94°	0.56	0.44	61.72°

3.3 Circular mean, resultant length and variance

Table 1 shows descriptives for the motor resonance data. For each group, the table contains a circular mean and a mean resultant length of the phase difference. The circular population mean, μ, indicates the average direction of a certain variable in the population. The population mean resultant length, ρ, is a statistic between 0 and 1 that gives us information on the spread of a circular variable in the population. It is interpreted as a precision measure where 0 means that the spread is large and 1 means that all data are concentrated at a single value. Sample statistics for these values are $\bar{\theta}$ for the mean and $\bar{R}$ for the mean resultant length.

Graphically we can illustrate the computation of $\bar{\theta}$ and $\bar{R}$ as shown in Figure 4. On the left side of this figure we see two sets of circular data. We represent a circular datapoint as a vector composed of the cosine and sine of the datapoint instead of one value a measure in degrees (or radians). E.g. for a score of 90° we have the following vector $(\cos(90°), \sin(90°))$. The solid vectors in Figure 4 each represent one circular datapoint. To compute the mean directions and resultant lengths for the datasets on the left we place the vectors head to toe, as in the right side of Figure 4. We then connect the toe of the first vector to the head of the last vector. This results in the dotted vectors on the right side of Figure 4. The direction of the dotted vector is the mean direction, $\bar{\theta}$ of the
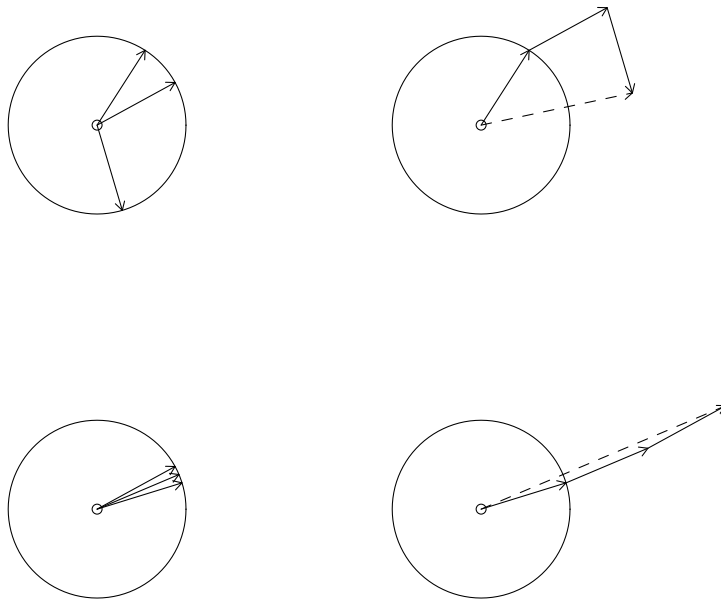
Figure 4: The computation of a circular mean and (mean) resultant length (right) for two sets of circular data (left). The solid lines are vectors representing the circular datapoints. The direction of the dotted vector is the mean direction and the length of the dotted vector is the resultant length.

vectors from which it was created. The length of the dotted vector is the resultant length. The mean resultant length, $\bar{R}$ is the length of this vector divided by the number of vectors from which it was created. In Figure 4 we see that the data in the bottom left figure are much more concentrated on the circle than the data in the upper left figure. This translates to the resultant length in the bottom right being larger than the resultant length (length of the dotted vector) in the upper right. Formulae for the computation of $\bar{\theta}$ and $\bar{R}$ can be found in [9].

In Table 1 we see that in the motor resonance data the circular mean of the phase difference for the explicit observation condition is highest with 49.55°. The mean phase differences for the semi-implicit and implicit observation conditions are lower at 18.45° and 31.94°. Moreover, the mean resultant lengths of the three groups differ. The phase differences of the individuals in the explicit observation condition are most concentrated with $\bar{R} = 0.77$. The phase differences differ more between the individuals in the semi-implicit and implicit observation conditions where the spread is larger at $\bar{R}$'s of 0.54 and 0.56 respectively.

Table 1 also shows a circular variance and standard deviation. The sample value, V_m for the circular variance is defined as $1 - \bar{R}$. Its interpretation is exactly opposite to the interpretation of the mean resultant length. A variance of one means that a variable has a very large spread and a variance of 0 means that all data are concentrated at one point. Note that unlike a linear variance the circular variance is bounded between 0 and 1. A sample circular standard deviation, v, can also be computed [9]. This deviation runs from 0 to infinity where higher values indicate a larger dispersion.

We have seen that both the average phase difference and variances of the phase difference seem to be different for the three conditions in the motor resonance data. To test whether these differences in circular means also exist in the population, we can use a a projected normal regression model. In the next section we will introduce this model and fit it to the motor resonance data.

4 A general linear model with a circular outcome

In this section we will introduce a projected normal circular regression model. This model falls within the embedding approach to circular data. Note that because it is a regression model we can also fit AN(C)OVA type models with it, we can thus refer to it as a circular GLM. As noted in the introduction, we choose to focus on a model from the embedding approach since this approach is quite flexible in the sense that it is relatively easy to translate models that exist for linear data to the circular case. We will first outline the theoretical background to the projected normal circular regression model and the embedding approach after which we will continue to fit an ANOVA to the motor resonance data. At the end of this section we will shortly consider two different methods for circular ANOVA.

4.1 The embedding approach to circular data

In the previous section, at the computation of the circular mean, we have seen that a circular variable θ, e.g. the phase difference in the motor resonance data, can be expressed as a unit vector $\boldsymbol{u}$ composed of the sine and the cosine of an angle $\boldsymbol{u} = (\cos\theta, \sin\theta)$. If we translate this to bivariate real space the cosine is the x-component and the sine is the y-component of an angle. In the embedding approach we assume that the circular variable $\boldsymbol{u}$ origins from the following relation:
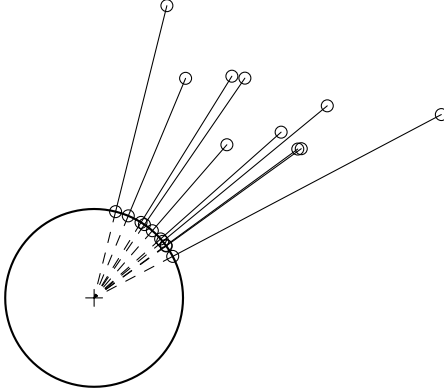
Figure 5: A set of datapoints in bivariate space $\boldsymbol{y}_i$ projected onto the circle to produce a set of datapoints $\boldsymbol{u}_i$. The lines connecting the datapoints to the origin of the circle represent r_i.

$$\boldsymbol{u} = \boldsymbol{y}/r \tag{1}$$

We assume that $\boldsymbol{u}$ can be multiplied by a positive linear variable r, $0 < r \leq \infty$, to obtain a bivariate normal variable $\boldsymbol{y}$. Note that in a real dataset with sample size n we have a set of vectors $\boldsymbol{u}_i$, one for each person $i = 1, \ldots, n$. This also means that there is an underlying vector $\boldsymbol{y}_i$ and value r_i for each person in our dataset. Figure 5 depicts this relation between $\boldsymbol{u}_i$ and $\boldsymbol{y}_i$.

The projected normal regression model is introduced by [16]. This model makes use of the embedding approach to circular data. The idea behind the embedding approach is to not conduct inference on the circular variable θ directly but to indirectly conduct inference on the underlying bivariate normal data in $\boldsymbol{y}$. This makes the embedding approach a flexible approach to circular data as many different and complex types of models already exist for bivariate normal data. However, both the vector $\boldsymbol{y}$ and value r are quantities that can not be obtained directly from θ. The estimation of these objects is a missing data problem. The solution for this missing data problem used in this paper relies on Bayesian estimation. More details on this Bayesian method are outlined by [17] and [18] for regression models.

Assuming that $\boldsymbol{y}$ has a bivariate normal distribution, the relation in (1) implies that θ has a projected normal distribution. The projected normal (PN) distribution for the cicular outcome θ is expressed as:

$$PN(\theta \mid \boldsymbol{\mu}, \boldsymbol{I}) = \frac{1}{2\pi} e^{-\frac{1}{2}||\boldsymbol{\mu}||^2} \left[1 + \frac{\boldsymbol{u}^t \boldsymbol{\mu} \Phi(\boldsymbol{u}^t \boldsymbol{\mu})}{\phi(\boldsymbol{u}^t \boldsymbol{\mu})} \right], \tag{2}$$

where θ is measured in radians $-\pi \leq \theta < \pi$, $\boldsymbol{\mu} = (\mu_1, \mu_2)^t \in \mathbb{R}^2$ is the mean vector of the distribution, the variance-covariance matrix $\boldsymbol{I}$ is an identity matrix, and $\boldsymbol{u}^t = (\cos\theta, \sin\theta)$. The terms $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cumulative distribution function and the probability density function of the standard

normal distribution respectively. When fitting a model using the PN distribution we are interested in modelling its mean vector $\boldsymbol{\mu}$.

In regression models, $\boldsymbol{\mu}$ has a regression structure. In a multiple regression model this is specified as follows:

$$\boldsymbol{\mu}_i = \begin{pmatrix} \mu_i^I \\ \mu_i^{II} \end{pmatrix} = \begin{pmatrix} (\boldsymbol{\beta}^I)^t \boldsymbol{x}_i^I \\ (\boldsymbol{\beta}^{II})^t \boldsymbol{x}_i^{II} \end{pmatrix}, \tag{3}$$

where I and II, the two components, are the x and y axis of the Cartesian plane, $i = 1, \ldots, n$, $\boldsymbol{x}_i$ is a vector of predictor values for individual i and each $\boldsymbol{\beta}$ is a vector with intercept and regression coefficients. To be able to estimate an intercept, the first component of $\boldsymbol{x}_i$ equals 1. Note that the vectors $\boldsymbol{x}_i$ are allowed to differ for the two components I and II. In the next section we will fit this type of model to the motor resonance data.

In terms of interpretation of the circular effect of a variable the two component structure in (3) poses a problem. The two components do not necessarily have a useful interpretation for each type of circular data, e.g. we cannot talk of a 12 o'clock and 6 o'clock axis in Figure 2. [18] outline how this problem can be overcome and introduce new interpretation tools in PN multiple regression models that transform the effects on the two components to an effect on the circle. In this tutorial we will use these tools and interpret coefficients of a circular effect and not the coefficients for each of the components.

4.2 Fitting an ANOVA model to the motor resonance data

In this section we will fit a circular ANOVA model to the motor resonance data using the package `bpnreg`. Note that the model from this package is in fact a regression model that we can speficy in such a way that it is mathematically equivalent to an ANOVA. First we will give the details on this model and subsequently we will discuss and interpret the results.

4.2.1 Fitting the model

To investigate the effect of condition on the phase difference we specify the prediction equation for the mean vector in the projected normal regression model as follows:

$$\boldsymbol{\mu}_i = \begin{pmatrix} \mu_i^I \\ \mu_i^{II} \end{pmatrix} = \begin{pmatrix} \beta_0^I + \beta_1^I \text{semi.implicit}_i + \beta_2^I \text{implicit}_i \\ \beta_0^{II} + \beta_1^{II} \text{semi.implicit}_i + \beta_2^{II} \text{implicit}_i \end{pmatrix}, \tag{4}$$

where the variables semi.implicit and implicit are dummy variables indicating condition membership, β_0^I and β_0^{II} are the intercepts and β_1^I, β_2^I, β_1^{II} and β_2^{II} are the regression coefficients of the model. Note that in this model we take the explicit observation condition as the reference condition. When we translate this to the ANOVA context, the intercepts, β_0^I and β_0^{II}, represent the mean for the explicit condition, $\beta_0^I + \beta_1^I$ and $\beta_0^{II} + \beta_1^{II}$ are expressions for the mean of the semi-implicit condition and $\beta_0^I + \beta_2^I$ and $\beta_0^{II} + \beta_2^{II}$ are expressions for the mean of the implicit condition.

We use the package `bpnreg` to fit the model. Because this is a Bayesian method we have to specify some parameters for the MCMC sampler that estimates the parameters of the model. MCMC
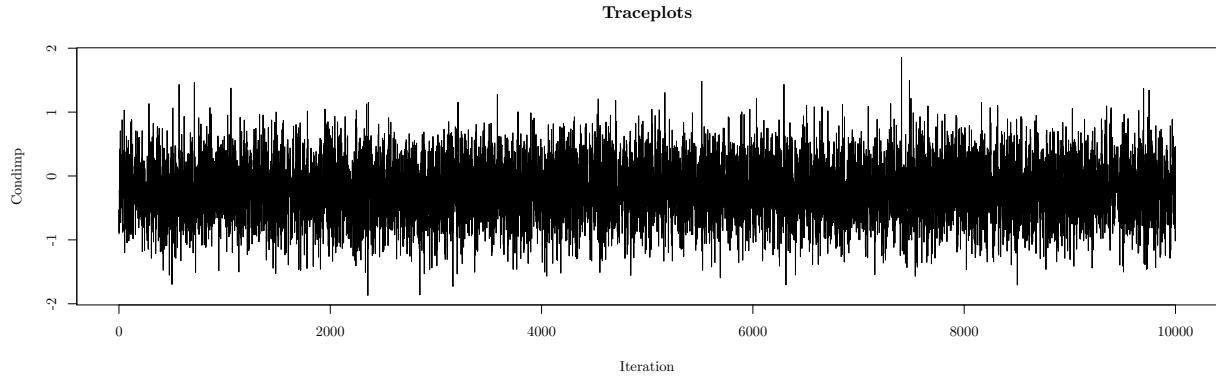
Figure 6: A traceplot showing convergence of the parameter β_2^I of the model fit to the motor resonance data.

methods are iterative and we thus need to specify the number of iterations that we want to run. We choose a relatively high number of the output iterations (`its = 10000`) to make sure that the sampler converges and that we don't need to run it again in case it did not. We choose the burn-in period to consist of 100 iterations (`burn = 100`). This means that we throw away the first 100 iterations to make sure that the iterations we keep are those at which the sampler has reached its equilibrium. We also choose the lag, that is how many iterations we want to keep, in this case every third iteration (`n.lag = 3`). We set a lag to prevent possible auto-correlation between the parameter estimates. In the next section we will elaborate further on how to check convergence and choose these MCMC parameters wisely. But first, we fit the model:

```
fit.Motor <- bpnr(pred.I = Phaserad ~ 1 + Cond, data = Motor,
                  its = 10000, burn = 100, n.lag = 3, seed = 101)
```

4.2.2 Convergence

In a Bayesian model that uses MCMC sampling for estimation we always have to assess convergence of the MCMC chain for all parameters in the model. A traceplot is one way to assess the convergence of a parameter. As an illustration we will show the traceplot for the MCMC chains for one of the parameters of the model.

From the previous traceplot we may conclude that the MCMC chain converged within 10000 iterations and a burn-in of 100. The plot looks like a 'fat-caterpillar' meaning that the chain has reached an equilibrium around a particular value. In case an MCMC chain does not converge we could have add more iterations, a larger `n.lag` or more burn-in iterations. We can also evaluate other convergence diagnostics. The focus of this paper however does not lie on Bayesian data analysis and therefore we refer to other works, e.g. [19], for more information on assessing convergence.

4.2.3 Results

To answer the question whether the conditions of the motor resonance data differ in their phase difference we investigate the estimated circular regression coefficients. In [18] it is explained how to

10

Table 2: Posterior estimates of the circular means of the phase diference for the three conditions of the motor resonance data.

Condition	Mode	Mean	sd	LB HPD	UB HPD
explicit	42.70°	45.56°	11.67°	22.26°	67.99°
semi-implicit	21.08°	19.40°	18.36°	-18.27°	55.22°
implicit	37.22°	33.47°	17.77°	-2.25°	68.22°

obtain circular regression coefficients for continuous variables and how to interpret them. However, because there are no continuous predictors in the model we want to estimate the circular means of the three conditions in the data. To do so we use (5) to transform the results from the two components of $\boldsymbol{\mu}$ to the circle.

$$
\begin{aligned}
\theta &= \operatorname{atan2}\left(\hat{y}^{II},\ \hat{y}^{I}\right) \\
&= \arctan\left(\frac{\hat{y}^{II}}{\hat{y}^{I}}\right) && \text{if} \quad \hat{y}^{I} > 0 \\
&= \arctan\left(\frac{\hat{y}^{II}}{\hat{y}^{I}}\right) + \pi && \text{if} \quad \hat{y}^{I} < 0 && \&\ \hat{y}^{II} \geq 0 \\
&= \arctan\left(\frac{\hat{y}^{II}}{\hat{y}^{I}}\right) - \pi && \text{if} \quad \hat{y}^{I} < 0 && \&\ \hat{y}^{II} < 0 \\
&= +\frac{\pi}{2} && \text{if} \quad \hat{y}^{I} = 0 && \&\ \hat{y}^{II} > 0 \\
&= -\frac{\pi}{2} && \text{if} \quad \hat{y}^{I} = 0 && \&\ \hat{y}^{II} < 0 \\
&= \text{undefined} && \text{if} \quad \hat{y}^{I} = 0 && \&\ \hat{y}^{II} = 0 && (5)
\end{aligned}
$$

E.g. for the circular mean of the semi-implicit condition $\hat{y}^{I} = \beta_0^{I} + \beta_1^{I}$ and $\hat{y}^{II} = \beta_0^{II} + \beta_1^{II}$ in (5). Because we use a Bayesian method we get the posterior distributions of the three circular means. Philosophically, in Bayesian statistics each parameter is said to have its own distribution. The posterior distribution is the result of the prior knowledge we have about a parameter before conducting a study, formalized as a 'prior' distribution (in this paper we choose non-informative priors for the parameters) and the information that lies in the data obtained from a study, formalized as the likelihood. The fact that we obtain the distribution of a parameter is convenient for inference purposes since this means that we do not just have a point estimate of a parameter (the mean or mode of the posterior distribution) but we also automatically get an uncertainty estimates (the standard deviation of the posterior distribution). For more background on Bayesian statistics see e.g. [19].

Summary statistics for the posterior distributions of the circular means for each condition are shown in Table 2. This table shows the posterior mean, mode, standard deviation (sd) and the lower and upper bound of the 95% highest posterior density interval (HPD). The standard deviation of a posterior is an estimate for the standard error of the parameter. The HPD interval is the smallest interval in which 95% of the posterior mass is located. In terms of interpretation, it is different from a frequentist confidence interval since HPD intervals allow for probability statements. For example, if the 95% HPD interval for a parameter μ runs from 2 to 4 we can say that the probability that μ

Table 3: Posterior estimates of the circular variances of the phase diference for the three conditions of the motor resonance data.

Condition	Mode	Mean	sd	LB HPD	UB HPD
explicit	0.21	0.26	0.09	0.09	0.45
semi-implicit	0.37	0.44	0.13	0.19	0.36
implicit	0.36	0.42	0.13	0.18	0.68

lies between 2 and 4 is 0.95.

HPD intervals can also be used to test whether a parameter is different from a certain value or whether two parameter estimates are different. In Table 2 we see that the HPD intervals of the circular means for the three conditions in the motor resonance data overlap. The circular mean of the phase difference is estimated at 47.70° (22.26°; 67.99°) for the explicit condition, 37.22° (-2.25°; 68.22°) for the implicit condition and 21.08° (-18.27°; 55.22°) for the semi-implicit condition. We thus conclude that the circular means for the three conditions do not differ and that there is no effect of condition on the average phase difference.

In addition to testing whether the circular means of the three conditions are different the circular ANOVA also allows us to test whether there is an effect of condition on the circular variances of the phase differences. Table 3 shows summary statistics for the posterior distributions of the circular variance for each condition. As expected the estimated circular variance for the explicit condition is lowest. However, the variances of the three groups do not differ; their HPD intervals overlap. We thus conclude that there if no effect of condition on the variance of the phase difference.

4.3   Other approaches to circular ANOVA

In the previous section we have tested whether the mean phase differences of the three conditions differ in the population using a Bayesian projected normal regression model. We can also do this using a frequentist ANOVA for circular data that tests the hypothesis $H_0 : \mu_{explicit} = \mu_{semi-implicit} = \mu_{implicit}$. One of such tests is the Watson-Williams test. This test can be perfomed using the function `watson.williams.test` in the `circular` package and is similar to an ANOVA for linear data interpretation-wise. As in the projected normal regression model we conclude from this test that the average phase differences of the three conditions are not significantly different: $F(2, 39) = 1.02$, $p > 0.05$.

As in ANOVA models for linear data, we have to meet a set of assumptions for this test to be valid. Firstly, the samples from the different conditions are assumed to be von-Mises distributed in the Watson-Williams test. Secondly, the samples are assumed to have the same circular variance. The von-Mises distribution is a type of distribution for circular data, it is unimodal with mean μ and concentration κ. The assumption of homogeneity of variance is tested within the `watson.williams.test` function itself. For the motor cognition data this assumption was met. The assumption of von-Misesness can be tested using e.g. the Watson's goodness of fit test for the von Mises distribution. If we perform this test on the phase differences of the three subgroups we conclude that only the phase differences from the `semi.implicit` and the `implicit` condition are von-Mises distributed (H_0 is not rejected). This means that it is not completely valid to perform the Watson-Williams test on the motor resonance data. In addition, the Watson-Williams test does not allow for the addition of covariates and thus cannot estimate AN(C)OVA models. Which the

PN circular GLM introduced in Section 4 does allow for.

Another method that we can use is a Bayesian circular GLM that falls within the intrinsic approach [20]. The intrinsic approach to circular data differs from the embedding approach in the sense that it uses distributions that are directly defined on the circle instead of projecting distributions in bivariate linear space onto the circle. The Watson-Williams test is thus also a method that falls within the intrinsic approach. The Bayesian circular GLM using the intrinsic approach can be fit using the package `circglmbayes`. When non-informative priors are used this methods will give the same results as the Watson-Williams test. However because it is a GLM this method can also fit ANCOVA and regression models. In contrast to the the method from [20], in which homogeneity of variances is assumed the PN circular GLM does not have this assumption. This allows for more flexible modelling including testing for an effect of predictor variables (e.g. condition in Section 4.2.3) on the variance.

5 Mixed-Effects models with a circular outcome

An additional advantage of employing the embedding approach to circular data over the intrinsic approach is that it is easier to model more complex data, e.g. repeated measures data, since we can 'borrow' methods to estimate these from the bivariate linear context. In this section we will introduce such a method: the circular mixed-effects model. We will first introduce a new dataset, the cognitive maps data, and give descriptive statistics. Then, we will shortly outline the theoretical background to the mixed-effects model and fit it to the cognitive maps data.

5.1 The cognitive maps data

The cognitive maps data is a subset of data from a study by [5] on the geometry of humans' knowledge of navigation space. In their study [5] amongst others conduct an experiment in which a total of 20 participants used virtual reality headsets to navigate through one of two virtual mazes. The navigation task consisted of walking from a start object to a target object. In a training phase they had learned to navigate between different pairs of start and target objects in one of two versions of the maze. The number of trials each participant completed in this training phase was recorded. In the test phase of the experiment participants first walked to a start object. When they had reached this object the maze dissappeared and only a "textured groundplane" of the maze remained visible. The participants then turned toward the location of the target object that they had remembered during the training phase and started to walk toward the target. The angular difference between the initial walking direction of a participant from the start object and the location of the target object, that is, the angular error, was recorded as an outcome variable in the experiment.

The type of maze is a between-subjects factor, participants either had to navigate through a 'Euclidean' maze or a 'non-Euclidean' maze. The Euclidean maze is the standard maze and is a maze just as we know it in the real world. The other version of the maze, the non-Euclidean maze, has exactly the same layout as the standard maze but it has virtual features that do not exist in reality. It namely contains wormholes by which participants can be 'teleported' from one place in the maze to another.

In the test phase of the experiment all participants had to complete 8 trials. In each of these trials participants had to walk to a specific target object. A within-subjects factor is the type of target

Table 4: Descriptives for the cognitive maps data with mean direction ($\bar{\theta}$) and mean resultant length ($\bar{R}$) of the angular error for each condition.

	Maze	Trial.type	$\bar{\theta}$	$\bar{R}$
Angular error	Euclidean	standard	-4.91°	0.89
		probe	4.46°	0.92
	non-Euclidean	standard	-17.59°	0.78
		probe	37.34°	0.93

object. Pairs of start and target objects were of two types: probe and standard. The probe objects were located near the entrance and exit of a wormhole in the non-Euclidean maze whereas the standard objects were located at some distance from the wormholes. For each of these two types of objects participants had to find 4 different targets resulting in a total of 8 trials per participant.

For this experiment we could be interested in the question whether the participants in the non-Euclidean maze make use of the wormholes when navigating to the target objects and whether this is true for both the probe and standard target objects. Due to the design of the mazes the expected angular error was larger if a participant used the wormhole to walk to the target object in the non-Euclidean maze. We can thus use the angular error, our outcome variable, to differentiate between participants that used the wormhole and those that took another path to the target object. Additionally we can control for the amount of trials that a participant completed in the training phase.

5.2 Descriptive Statistics

The cognitive maps data is incorporated in the package `bpnreg` as the dataframe `Maps`. This dataframe has 160 rows; there are 20 subjects that each completed 8 trials. The data contains an index variable for the subject `Subject` (N = 20) and trial number `Trial.no` (n = 8). It also includes variables indicating the type of maze `Maze`, a between-subjects factor, and type of trial `Trial.type`, a within-subjects factor. The variable `Learn` indicates the amount of learning trials completed. `L.c` is a centered version of this variable. The angular error is contained in the variables `Error` and `Error.rad` in degrees and radians respectively. Descriptives for this data are shown in Table 4. Note that we averaged over subjects and the trials of each type. The circular mean of the angular error for the standard trials in the Euclidean maze is thus an average over 10 participants and 4 trials. We see what the angular errors of both trial types for the non-Euclidean maze deviate more from 0° (direction of the target object) than for the Euclidean maze.

5.3 Fitting a mixed-effects model to the cognitive maps data

In this section we will first introduce a circular mixed effects model and fit this model to the cognitive maps data. Next we discuss the output produced by the `bpnreg` package. We will discuss the interpretation of fixed and random effects and model fit.

5.3.1 The embedding approach for mixed-effects models

The circular mixed-effects model from the package `bpnreg` is also based on the embedding approach to circular data. The basic idea behind this approach is the same as outlined in Section 4.1. In a real dataset we have a set of outcome vectors $\boldsymbol{u}_{ij}$, one for each measurement j within a higher level observation i. The solution for the missing data problem in the estimation of $\boldsymbol{y}_{ij}$ and r_{ij} is outlined in [21].

For the cognitive maps data with $i = 1, \ldots, 20$ individuals and $j = 1, \ldots, 8$ measurements per individual we fit a mixed-effects model to investigate the influence of the type of Maze, type of trial and amount of learning trials on the angular error. The prediction for the mean vector in this model, $\boldsymbol{\mu}$, is specified as follows:

$$\boldsymbol{\mu}_{ij} = \begin{pmatrix} \mu_{ij}^I \\ \mu_{ij}^{II} \end{pmatrix} = \begin{pmatrix} \beta_0^I + \beta_1^I \text{Maze}_i + \beta_2^I \text{Trial.type}_{ij} + \beta_3^I \text{L.c}_i + b_{0i}^I \\ \beta_0^{II} + \beta_1^{II} \text{Maze}_i + \beta_2^{II} \text{Trial.type}_{ij} + \beta_3^{II} \text{L.c}_i + b_{0i}^{II} \end{pmatrix}, \tag{6}$$

where the variables `Maze` and `Trial.type` are dummy variables, β_0^I and β_0^{II} are the fixed intercepts, b_{0i}^I and b_{0i}^{II} are the random intercepts and β_1^I, β_2^I, β_3^I, β_1^{II}, β_2^{II} and β_3^{II} are the fixed regression coefficients of the model. Note that in this model we take the Euclidean maze and standard trials as reference conditions.

The interpretation problems caused by the two component structure in (6) is of a similar nature as the one in the GLM model. [22] introduce new tools that solve the interpretation of circular effects in PN mixed-effects models. In this tutorial we will also use these tools.

5.3.2 Fitting the model

To fit the model in (6) we use the `bpnme()` function from the package `bpnreg`. We also need to specify some parameters for the MCMC sampler that estimates the model. We specify the output iterations (10000), the amount of burn-in (100) and how many iterations we want to keep (`n.lag = 3`). Convergence was checked in the same manner as for the regression model in the previous section and was reached using the settings for the MCMC algorithm we just specified.

```
fit.Maps <- bpnme(pred.I = Error.rad ~ Maze + Trial.type + L.c + (1|Subject),
                  data = Maps,
                  its = 10000, burn = 1000, n.lag = 3, seed = 101)
```

Note that the syntax for the model specification in this function is similar to that of the package `lme4` for fitting (non-circular) mixed-effects models.

5.3.3 Fixed Effects

Next we investigate the coefficients of the fixed effects for this model. First we show results for the categorical variables type of maze (`Maze`) and type of trial (`Trial.type`).

Categorical variables

Table 5: Posterior estimates of the circular mean of the angular error for each condition.

	Maze	Trial.type	mode	mean	sd	LB HPD	UB HPD
Angular error	Euclidean	standard	-12.97°	-13.48°	3.9°	-21.42°	-6.06°
		probe	11.38°	11.78°	3.29°	5.26°	18.30°
	non-Euclidean	standard	-1.42°	-2.04°	6.68°	-15.75°	10.49°
		probe	31.04°	30.37°	4.31°	22.03°	38.92°

Table 6: Posterior estimates of the coefficients of the effect of L.c on the angular error.

Coefficient	mode	mean	sd	LB HPD	UB HPD
b_c	-0.89°	-0.21°	1.73°	-2.84°	2.55°
SAM	0.58°	-0.84°	90.80°	-11.51°	12.73°
AS	-0.63°	-1.11°	92.44°	-13.17°	13.14°

Table 5 shows summary statistics of the posterior of the average angular error for each of the categories. Note that because there is a continuous predictor in the model the posterior estimates represent a marginal effect, they are the effect for an individual with a 0 score on the continuous predictor `L.c`. Because we centered this predictor this means that this is the effect for an individual that has completed an average number of training trials.

By looking at the 95% HPD intervals of the angular errors in Table 5 we can test whether the type of maze and type of trial on average has an influence on the angular error and thus whether participants make use of the wormhole. For the standard trials we see that the HPD intervals of the angular error in the Euclidean and non-Euclidean overlaps and that thus the angular error is not different. This means that in the standard trials the participants on average did not make use of the wormholes in the non-Euclidean maze. For the probe trials however, the HPD intervals of the Euclidean and non-Euclidean do not overlap and thus the angular error is different. This means that in the probe trials, the participants on average did make use of the wormholes in the non-Euclidean maze.

Continuous variables

For the continuous variable `L.c` we get a set of parameters, b_c, SAM and AS, describing its effect on the circle. How these parameters are computed is described in [18] and [22]. In this paper we will only focus on how to interpret them.

In Figure 7 a circular regression line for the effect of a predictor x on the circular outcome is shown. Because the outcome variable is measured on a circular scale, the slope of this line (the effect of x) is not constant but different for different x values. The regression line can be described using the three circular coefficients b_c, SAM and AS. The coefficient b_c represents the slope of the circular regression line at the inflection point (the square in Figure 7). However, this may not be a representative effect for each dataset as the inflection point can lie in the extremes of the data (as in Figure 7) or even completely outside the range of the predictor x. Therefore two additional circular coefficients were developed by [18], the slope at the mean SAM and the average slope AS. The coefficient SAM represents the slope of the circular regression line at the average of the predictor ($\bar{x}$) and the coefficients AS represents the average slope over all values of x.

For the effect of L.c on the angular error in the cognitive maps data, the HPD intervals for all three circular coefficients, b_c, SAM and AS include 0 (see Table 6. Thus we conclude that at the
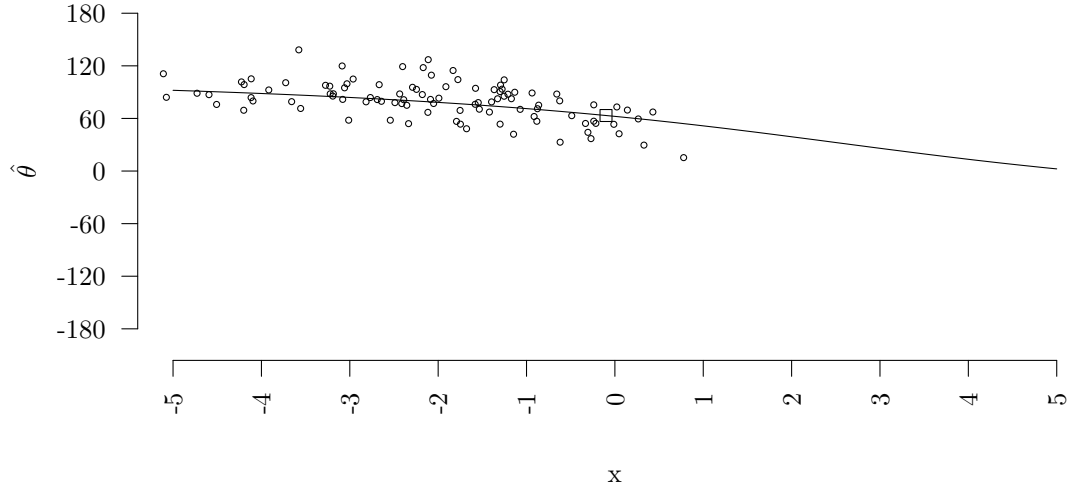
Figure 7: Predicted circular regression line for the relation between a linear predictor x and a predicted circular outcome θ together with the original datapoints. The square indicates the inflection point of the regression line.

inflection point, at the average predictor value and on average the number of training trials (`L.c`) does not influence the angular error. This is in fact a good thing, since it shows that the training phase of the experiment worked to get all participants at the same level. For educational purposes we will however continue to give the interpretation of the coefficients. The SAM is interpreted as follows: at the average `L.c`, for a 1 unit increase in `L.c` the angular error increases with 0.58 degrees. The AS can be interpreted as: on average, for a 1 unit increase in `L.c` the angular error decreases with 0.63 degrees. The b_c can be interpreted as: at the inflection point, for a 1 unit increase in `L.c` the angular error decreases with 0.89 degrees.

5.3.4 Random Effects

In mixed-effects models we are also interested in evaluating the variance of the random effects. In the model for the cognitive maps data we included a random intercept. This means that we estimate a separate intercept for each participant. How to compute random effect variances on the circle is outlined in [22]. For the cognitive maps data the posterior mode of the intercept variance on the circle is estimated at $3.5 * 10^{-5}$ and its HPD interval is $(4.2 * 10^{-6}; 1.4 * 10^{-3})$. This variance is very low which means that the participants do not differ a lot in their individual intercept estimates. Note that this is not necessarily problematic. In some cases we are not interested in the variances of the random efects but simply want to fit a mixed-effects model because we have within factors, such as `Trial.type`, that cannot be properly incorporated in a standard regression model.

17

5.3.5 Model Comparison

When fitting mixed-effects (or multilevel) models we often fit a set of nested models to our data and follow a model building strategy [23]. We do this in case we have no specific model in mind that we want to test and want to explore the individual contributions of variables or groups of variables to the model. Such a model building strategy can be done top-down, starting with the most complex model, or bottom-up, starting with the simplest model. Here we use a bottom-up strategy and start with the so called intercept-only model, a model containing only a fixed and random intercept:

```
fit.MapsIO <- bpnme(pred.I = Error.rad ~ (1|Subject),
                    data = Maps,
                    its = 10000, burn = 1000, n.lag = 3, seed = 101)
```

We then update this model with fixed effects for the predictors at the lowest level (within-subjects factors), in this case `Trial.type`. We do this to check whether the set of predictors improved the fit of the model and can explain a part of the random intercept variance from the intercept-only model.

```
fit.Maps1p <- bpnme(pred.I = Error.rad ~ Trial.type + (1|Subject),
                    data = Maps,
                    its = 10000, burn = 1000, n.lag = 3, seed = 101)
```

We then add fixed effects for the predictors at the higher level (between-subjects factors), in this case `Maze` and `L.c`. Again we do this to check whether they improve the fit of the model and whether they can explain a part of the random intercept variance.

```
fit.Maps <- bpnme(pred.I = Error.rad ~ Maze + Trial.type + L.c + (1|Subject),
                  data = Maps,
                  its = 10000, burn = 1000, n.lag = 3, seed = 101)
```

Because we have already seen that the effect of `L.c` was not different from 0 we also fit the model with only the `Maze` and `Trial.type` predictors.

```
fit.Maps2 <- bpnme(pred.I = Error.rad ~ Maze + Trial.type + (1|Subject),
                   data = Maps,
                   its = 10000, burn = 1000, n.lag = 3, seed = 101)
```

Additional steps, such as adding random slopes for first level predictors and cross-level interactions, can be taken. In this paper we will however restrict the analysis to the previous three models.

Model fit

To assess the fit of the models we look at 4 different model fit criteria: two version of the deviance information criterion (DIC and DIC_{alt}) and two versions of the Watanabe-Akaike information criterion ($WAIC_1$ and $WAIC_2$). We choose these four criteria because they are specifically useful in Bayesian models where MCMC methods have been used to estimate the parameters. All four criteria have a fit part consisting of a measure based on the loglikelihood and include a penalty in the form of an effective number of parameters. For all criteria lower values indicate better fit. [19] describes how to compute these criteria. Table 7 shows the results of these criteria for 4 different models.

In the results for the example we see that the fit improves in all 4 model diagnostics for each model except for the last one. This means that the predictor `Trial.type` improves the fit of the model

Table 7: Model fit criteria for several models fit to the cognitive maps data.

Criterion	Intercept-only	Trial.type	Trial.type + Maze	Trial.type + Maze + L.c
DIC	304.61	267.91	253.97	257.94
DIC_{alt}	324.33	286.97	257.14	260.78
$WAIC_1$	308.41	271.61	255.00	258.41
$WAIC_2$	308.43	271.77	255.40	259.02

over the intercept-only model and that the predictors `Maze` and `Trial.type` together improve the fit of the model over the model with only the `Trial.type` predictor. Because the variable `L.c` had no effect it is as expected that this predictor does not improve the fit of the most right model over the model with the `Maze` and `Trial.type` predictors. We conclude that the model with the predictors `Trial.type` and `Maze` fits best.

Explained variance

Apart from information about whether adding predictors improves the fit of the model we are also interested in whether these predictors explain a part of the random effect variances. For the cognitive maps data we are interested in whether the `Maze` and `Trial.type` predictors explain a part of the variance in individual intercepts. To assess this we compare the posterior estimates of the circular random intercept for the intercept-only model and the model with the `Maze` and `Trial.type` predictors.

The posterior mode of the intercept variance in the intercept-only model equals $6.61 * 10^{-5} (8.20 * 10^{-6}; 3.62 * 10^{-3})$, the posterior mode of the intercept variance in the model with the `Maze` and `Trial.type` predictors equals $3.25 * 10^{-5} (4.40 * 10^{-6}; 1.59 * 10^{-3})$. First of all, note that in the intercept-only model there is almost no random intercept variance. The posterior mode of the circular variance is very close to 0. Compared to the estimates of the variance in the intercept-only model there is hardly any change in estimates for the variance in the mode with `Maze` and `Trial.type`. Furthermore, their HPD intervals have a very large overlap. We thus conclude that the variables `Maze` and `Trial.type` did not explain any variance in the random intercepts. This makes sense as there was harly any intercept variance in the intercept-only model to begin with.

6   Concluding remarks

In this paper we have given a tutorial for researchers in cognitive psychology on how to analyse circular data using the package `bpnreg`. We have covered data inspection in Section 3, the fitting of a Bayesian circular GLM in Section 4 and the fitting of a Bayesain mixed-effects model in Section 5. We have also given a short introduction into the theoretical background of these models in Section 4.1 and 5.3.1.

Apart from the embedding approach to circular data, as used in this tutorial, there are two other approaches to the analysis of circular data. In the wrapping approach the data on the circle is assumed to have origined from wrapping a univariate distribution on the real line onto the circle. In the intrinsic approach distributions, such as the von Mises distribution, are directly defined on the circle. For both approaches models have been described in the literature [20,24–27]. The regression model using the intrinsic approach from [27] is implemented in the package `circular` and the circular general linear model from [20] is implemented in the package `circglmmbayes`. For

neither approach however a detailed tutorial describing how to analyze circular data using the functions from their package has been written thus far. Furthermore, the PN approach to circular modelling has the additional advantage that it is relatively easy to fit more complex models, e.g. the mixed-effects model in this tutorial.

References

1. Rivest L-P, Duchesne T, Nicosia A, Fortin D. A general angular regression model for the analysis of data on animal movement in ecology. Journal of the Royal Statistical Society: Series C (Applied Statistics). 2015;63: 445–463. doi:10.1111/rssc.12124

2. Mardia KV, Taylor CC, Subramaniam GK. Protein bioinformatics and mixtures of bivariate von mises distributions for angular data. Biometrics. 2006;63: 505–512. doi:10.1111/j.1541-0420.2006.00682.x

3. Rutishauser U, Ross IB, Mamelak AN, Schuman EM. Human memory strength is predicted by theta-frequency phase-locking of single neurons. Nature. 2010;464: 903–907. doi:10.1038/nature08860

4. Brunyé TT, Burte H, Houck LA, Taylor HA. The map in our head is not oriented north: Evidence from a real-world environment. PLoS ONE. 2015;10: 1–12. doi:10.1371/journal.pone.0135803

5. Warren WH, Rothman DB, Schnapp BH, Ericson JD. Wormholes in virtual space: From cognitive maps to cognitive graphs. Cognition. 2017;166: 152–163. doi:10.1016/j.cognition.2017.05.020

6. Matsushima EH, Vaz AM, Cazuza RA, Ribeiro Filho NP. Independence of egocentric and exocentric direction processing in visual space. Psychology & Neuroscience. 2014;7: 277–284. doi:10.3922/j.psns.2014.050

7. Heyes SB, Zokaei N, Husain M. Longitudinal development of visual working memory precision in childhood and early adolescence. Cognitive Development. 2016;39: 36–44. doi:10.1016/j.cogdev.2016.03.004

8. Kirschner S, Tomasello M. Joint drumming: Social context facilitates synchronization in preschool children. Journal of experimental child psychology. 2009;102: 299–314. doi:10.1016/j.jecp.2008.07.005

9. Fisher NI. Statistical analysis of circular data. Cambridge: Cambridge University Press; 1995.

10. Batschelet E. Circular statistics in biology. London: Academic Press; 1981.

11. Pewsey A, Neuhäuser M, Ruxton GD. Circular statistics in r. Oxford University Press; 2013.

12. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2017. Available: https://www.R-project.org/

13. Cremers J. Bpnreg: Bayesian projected normal regression models for circular data [Internet]. 2018. Available: https://CRAN.R-project.org/package=bpnreg

14. Agostinelli C, Lund U. R package `circular`: Circular statistics (version 0.4-93) [Internet]. 2017. Available: https://r-forge.r-project.org/projects/circular/

15. Puglisi G, Leonetti A, Landau A, Fornia L, Cerri G, Borroni P. The role of attention in human motor resonance. PloS one. 2017;12: e0177457. doi:10.1371/journal.pone.0177457

16. Presnell B, Morrison SP, Littell RC. Projected multivariate linear models for directional data. Journal of the American Statistical Association. Taylor & Francis Group; 1998;93: 1068–1077. doi:10.1080/01621459.1998.10473768

17. Nuñez-Antonio G, Gutiérrez-Peña E, Escarela G. A Bayesian regression model for circular data based on the projected normal distribution. Statistical Modelling. SAGE Publications; 2011;11: 185–201. doi:10.1177/1471082X1001100301

18. Cremers J, Mulder KT, Klugkist I. Circular interpretation of regression coefficients. British Journal of Mathematical and Statistical Psychology. 2018;71: 75–95. doi:10.1111/bmsp.12108

19. Gelman A, Carlin J, Stern H, Dunson D, Vehtari A, Rubin D. Bayesian data analysis. 3rd ed. Boca Raton, FL: Chapman & Hall/CRC; 2014.

20. Mulder K, Klugkist I. Bayesian estimation and hypothesis tests for a circular generalized linear model. Journal of Mathematical Psychology. 2017;80: 4–14. doi:10.1016/j.jmp.2017.07.001

21. Nuñez-Antonio G, Gutiérrez-Peña E. A Bayesian model for longitudinal circular data based on the projected normal distribution. Computational Statistics & Data Analysis. Elsevier; 2014;71: 506–519. doi:10.1016/j.csda.2012.07.025

22. Cremers J, Pennings HJM, Mainhard MT, Klugkist I. Longitudinal circular modelling of circumplex measurements for teacher behavior. Working Paper.

23. Hox JJ. Multilevel analysis: Techniques and applications. Hove: Routledge; 2002.

24. Ravindran P, Ghosh SK. Bayesian analysis of circular data using wrapped distributions. Journal of Statistical Theory and Practice. Taylor & Francis; 2011;5: 547–561. doi:10.1080/15598608.2011.10483731

25. Lagona F. Regression analysis of correlated circular data based on the multivariate von mises distribution. Environmental and ecological statistics. Springer; 2016;23: 89–113. doi:10.1007/s10651-015-0330-y

26. Gill J, Hangartner D. Circular data in political science and how to handle it. Political Analysis. SPM-PMSAPSA; 2010;18: 316–336. doi:10.1093/pan/mpq009

27. Fisher NI, Lee AJ. Regression models for an angular response. Biometrics. JSTOR; 1992;48: 665–677.