

**Evaluating Informative Hypotheses with Equality and Inequality Constraints:
A Tutorial Using the Bayes Factor via the Encompassing Prior Approach**

Claudio Zandonella Callegher¹, Tatiana Marci¹, Pietro De Carli², and Gianmarco Altoè¹


¹Department of Developmental Psychology and Socialization - University of Padova


²Department of Psychology - University of Milano-Bicocca

Author Note

Claudio Zandonella Callegher  <https://orcid.org/0000-0001-7721-6318>

Tatiana Marci  <https://orcid.org/0000-0002-2813-0312>

Pietro De Carli  <https://orcid.org/0000-0001-6538-2175>

Gianmarco Altoè  <https://orcid.org/0000-0003-1154-9528>

All materials and data are available at <https://doi.org/10.17605/OSF.IO/JKMDF>, GitHub repository <https://github.com/ClaudioZandonella/Attachment>. Online supplemental material available via GitHub Pages at <https://claudiozandonella.github.io/Attachment/>.

Correspondence concerning this article should be addressed to Claudio Zandonella Callegher, Department of Developmental Psychology and Socialisation, University of Padua, Via Venezia, 8 - 35131 Padova, Italy.

E-mail: claudiozandonella@gmail.com

Abstract

When conducting a study, researchers usually have expectations based on hypotheses or theoretical perspectives they want to evaluate. Equality and inequality constraints on the model parameters are used to formalize researchers' expectations or theoretical perspectives into the so-called informative hypotheses. However, traditional statistical approaches, such as the Null Hypothesis Significance Testing (NHST) or the model comparison using information criteria (e.g., AIC and BIC), are unsuitable for testing complex informative hypotheses. An alternative approach is to use the Bayes factor. In particular, the Bayes factor based on the encompassing prior approach allows researchers to easily evaluate complex informative hypotheses in a wide range of statistical models (e.g., generalized linear). This paper provides a detailed introduction to the Bayes factor with encompassing prior. First, all steps and elements involved in the formalization of informative hypotheses and the computation of the Bayes factor using the encompassing prior approach are described. Next, we apply this method to a real-case scenario, considering the attachment theory. Specifically, we analyzed the relative influence of maternal and paternal attachment on children's social-emotional development by comparing the various theoretical perspectives debated in the literature.

Keywords: Bayes Factor, Encompassing Prior, Informative Hypothesis, Equality and Inequality Constraints, Attachment Theory

Evaluating Informative Hypotheses with Equality and Inequality Constraints: A Tutorial Using the Bayes Factor via the Encompassing Prior Approach

1 Introduction

When conducting a study, researchers usually have expectations based on hypotheses or theoretical perspectives they want to evaluate according to the observed data. In fact, the evaluation of research and theoretical hypotheses is one of the principal goals of empirical research.

The dominant statistical approach to evaluate research hypotheses in psychology is the Null Hypothesis Significance Testing (NHST). In the literature, however, the NHST's utility and validity are largely debated (Wasserstein et al., 2019). This approach presents several limitations. First, the NHST places a narrow focus on statistical testing rather than on the formalization of hypotheses, leading researchers to evaluate data against the catch-all null hypothesis that nothing is going on rather than testing their specific expectation (and the alternative hypothesis is rarely formalized). Second, the p -value does not quantify the evidence in favor of one hypothesis; therefore, it is not possible to “*accept*” a hypothesis but only to “*reject*” it. This situation has some drawbacks because nonsignificant results leave researchers in a state of indecision. Third, the NHST does not allow researchers to test multiple hypotheses at the same time. With the NHST, the null hypothesis is tested only against a single alternative. Fourth, the NHST is unsuitable for testing broad classes of hypotheses with equality and inequality constraints (Mulder & Olsson-Collentine, 2019). Researchers can evaluate expectations using one-side tests; however, when more groups or more variables are involved, it is not possible to evaluate complex parameter constraints that reflect researchers' expectations (van de Schoot et al., 2011).

Model comparison is a different approach that allows researchers to compare multiple hypotheses and identify the one most supported by the data (McElreath, 2020). Researchers first formalize hypotheses as statistical models according to their expectations

or theoretical perspectives. Therefore, it is possible to determine which is the most supported model among those considered according to the data. To do so, a popular approach is to use information criteria, such as the AIC or BIC criteria, that estimate a model's ability to predict new data (Akaike, 1973; Schwarz, 1978; Wagenmakers & Farrell, 2004). Model comparison has several advantages; in particular, it allows researchers to directly compute each model's relative plausibility given the data and the set of models considered. However, model comparison using traditional information criteria (i.e., AIC and BIC) is not appropriate to evaluate hypotheses that include complex parameter constraints reflecting researchers' expectations. In fact, traditional information criteria penalize models' complexity according to the number of parameters, but they do not take into account possible order constraints on the parameters. Modifications of the AIC have been proposed to account for parameters order restrictions, leading to the order-restricted information criterion (ORIC; Anraku, 1999) and its recent generalization into the so-called generalized order-restricted information criterion approximation (GORICA; Altinisik et al., 2021; R. M. Kuiper et al., 2011, 2012). The GORICA allows to evaluate (in)equality constrained hypotheses under a very broad range of statistical models (Altinisik et al., 2021; R. Kuiper, 2022), including generalized linear models (GLMs), generalized linear mixed models (GLMMs), and structural equation models (SEMs). Consider also Mulder and Raftery (2022) for an extension of the BIC for evaluating order-constrained models.

An alternative criterium to evaluate research hypotheses in a model comparison is the Bayes factor. During the last 25 years, interest in the Bayes factor has increased, and its use has been proposed as the solution to the NHST's critical issues (Heck et al., 2020; Mulder & Wagenmakers, 2016). Although it has its own limitations and it does not solve the fundamental issues of the misuse of statistical techniques (Gelman et al., 2013; Schad, Nicenboim, et al., 2021), the Bayes factor offers some clear advantages. In particular, it allows us to compare hypotheses obtaining a relative index of evidence, like the information criteria, but it also allows us to easily compare complex research hypotheses. In fact,

researchers can formalize the so-called “*informative hypotheses*”, hypotheses containing information about the ordering of the model parameters, according to their expectations or theoretical perspectives, using equality and inequality constraints (van de Schoot et al., 2011). Researchers can then compare these hypotheses against each other, using the Bayes factor.

The evaluation of informative hypotheses is of particular interest to researchers, as it allows them to assess specific complex expectations and theoretical perspectives directly. In the literature, a particular approach allowing researchers to easily compute the Bayes factor with complex informative hypotheses has received increasing attention: “*the Bayes factor with encompassing prior*”. van de Schoot et al. (2011) presented a general introduction to informative hypothesis testing using the Bayes factor with the encompassing prior approach whereas Hoijtink (2012), Klugkist and Hoijtink (2007), and Klugkist, Kato, et al. (2005) offered a more detailed description of this method. In other studies, researchers considered the application of this approach with specific statistical models, such as analysis of variance models (Klugkist, Laudy, et al., 2005), mixed-effect models (Kato & Hoijtink, 2006), evaluation of correlation coefficients (Mulder, 2016), and multinomial models (Heck & Davis-Stober, 2019). In addition, Gu et al. (2014) proposed a general approximate procedure to evaluate inequality constraints in a wide range of statistical models.

However, in all the studies mentioned above, researchers considered only informative hypotheses with inequality constraints. To evaluate equality and inequality constraints in the same hypothesis, researchers had to approximate equality constraints to “*about equality constraints*” (i.e., a equality constraints of type $\theta_i = \theta_j$ is approximated to $|\theta_i - \theta_j| < \xi$ for a value of ξ small enough).

Only recently, Gu et al. (2018) introduced an approximate procedure to evaluate equality and inequality constraints, using the Bayes factor with the encompassing prior approach in a wide range of statistical models (i.e., GLMs, GLMMs, and SEMs). Subsequently, Mulder and Gelissen (2019) extended this approach to generalized

multivariate probit models, and, finally, Mulder and Olsson-Collentine (2019) proposed an accurate procedure (i.e., not based on an approximation) that allows for testing of informative hypotheses with equality and inequality constraints in linear regression models.

The development and implementation of these methods are of particular interest because although they have some limits, they allow researchers to directly evaluate their expectations and hypotheses. The available literature, however, is rather technical. The complexity of these articles makes it difficult for researchers not familiar with this approach to clearly understand all the steps involved. On the other hand, articles offering a general introduction to the Bayes factor do not provide enough details to allow readers to autonomously apply this method to their problems; instead, they usually rely on ad hoc solutions implemented in some statistical software. For example, Hoijtink et al. (2019) offered an excellent introduction to the Bayes factor for the testing hypotheses discussing all the main aspects. However, the description of the evaluation of competing informative hypotheses is limited and no specific details are provided regarding the adopted encompassing prior approach. Moreover, using ad-hoc solutions that automate the analysis, such as the functions available in the R packages `bain` (Gu et al., 2021) and `BFpack` (Mulder et al., 2021), prevent the users from fully understanding all the steps and limits entailed in this approach.

The aim of this paper, therefore, is to offer a clear and detailed description of the Bayes factor with the encompassing prior approach to allow other researchers to consciously apply this approach in their studies. In particular, we refer to the approximated method Gu et al. (2018) proposed, as it applies to a wider range of conditions than the more accurate approach Mulder and Olsson-Collentine (2019) introduced. In particular, researchers can use the proposed method in the case of generalized linear mixed models and structural equation models, two very general families of statistical models. Moreover, by providing an exhaustive description of all of the steps of this approach, researchers do not have to rely on already available solutions; instead, they

can implement this approach according to their specific needs, aware of its strengths and limits. The paper is organized as follows. First, we introduce the method, providing a detailed description of all steps and elements involved in the formalization of informative hypotheses and Bayes factor computation. We then present an application of the method to real data to discuss the typical issues encountered in real complex scenarios.

2 Bayes Factor for Informative Hypothesis Testing

The evaluation of informative hypotheses with equality and/or inequality constraints involves several steps and elements. In this section, first, we describe how to formulate informative hypotheses. We then introduce the Bayes factor, considering the encompassing prior approach based on the approximated method proposed by Gu et al. (2018).

2.1 Formulation of Informative Hypotheses

Researchers can define informative hypotheses according to their expectation, evidence from the literature, or theoretical perspectives, and they form them with equality and/or inequality constraints on certain model parameters. These constraints are obtained as a linear combination of certain parameters and eventual constant values. For example, it is possible to state that two parameters are equal ($\theta_i = \theta_j$) or that one parameter is greater than another ($\theta_i > \theta_j$) or to express other complex conditions, such as the difference between two parameter is less than a given value ($\theta_i - \theta_j < .5$; $2 \times \theta_i - \theta_j < .1 \times \sigma$).

Therefore, an informative hypothesis, H_i , with equality and inequality constraints can be expressed in the form

$$H_i : R_E \boldsymbol{\theta} = r_E \quad \& \quad R_I \boldsymbol{\theta} > r_I, \quad (1)$$

where R_E is a matrix expressing the equality constraints and r_E is a vector containing the equality constraints' constant values. R_I is a matrix expressing the inequality constraints, and r_I is a vector containing the inequality constraints' constant values. Finally, $\boldsymbol{\theta}$ is a vector with the model parameters involved in the constraints.

As an example, consider a study evaluating the efficacy of a new psychological

treatment that is supposed to improve a given cognitive ability. In the study, a control group receiving no treatment and another group receiving the traditional treatment were included as a comparison. Moreover, imagine that the new psychological treatment was administered in three different modalities to different groups: individually, in-group, and online. Researchers expect no differences between the three modalities, but they hypothesize that the new treatment will perform better than the traditional one, and this result, in turn, will be better than the no-treatment control condition. We can express this hypothesis as

$$H_i : \theta_{control} < \theta_{traditional} < \theta_{individual} = \theta_{group} = \theta_{online}, \quad (2)$$

were $\theta_{control}$ is the parameter estimating the average score of the no-treatment control group, $\theta_{traditional}$ is the parameter estimating the average score of the group receiving the traditional treatment, and $\theta_{individual}$, θ_{group} , and θ_{online} are the parameters estimating the average scores of the groups receiving the new treatment individually, in-group, and online, respectively. The corresponding formulation of the hypothesis using the matrix notation

previously introduced is

$$H_i : \quad R_E \boldsymbol{\theta} = \begin{bmatrix} 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \theta_{control} \\ \theta_{traditional} \\ \theta_{individual} \\ \theta_{group} \\ \theta_{online} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = r_E, \quad (3)$$

$$R_I \boldsymbol{\theta} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \theta_{control} \\ \theta_{traditional} \\ \theta_{individual} \\ \theta_{group} \\ \theta_{online} \end{bmatrix} > \begin{bmatrix} 0 \\ 0 \end{bmatrix} = r_I.$$

Note how each row of R_E and R_I matrices expresses an equality or an inequality constraint, respectively. For example, in the first row of R_E , we have $\theta_{individual} - \theta_{group} = 0$ (i.e., $\theta_{individual} = \theta_{group}$) and in the first row of R_I we have $\theta_{traditional} - \theta_{control} > 0$ (i.e., $\theta_{traditional} > \theta_{control}$).

Now that we have explained how to define an informative hypothesis with equality and inequality constraints introducing an appropriate notation, let's see how to evaluate an informative hypothesis using the Bayes factor with the encompassing prior approach.

2.2 Bayes Factor

The Bayes factor of hypothesis H_1 against a competing hypothesis, H_2 , is defined as the ratio between the marginal likelihoods of the two hypotheses:

$$BF_{12} = \frac{Pr(Y|H_1)}{Pr(Y|H_2)} = \frac{\int l(Y|\theta_1, H_1) \pi(\theta_1|H_1) d\theta_1}{\int l(Y|\theta_2, H_2) \pi(\theta_2|H_2) d\theta_2}, \quad (4)$$

where Y indicates the data, θ_i is the vector of parameters under the hypothesis H_i ,

$l(Y|\theta_i, H_i)$ is the likelihood function under the hypothesis H_i , and $\pi(\theta_i|H_i)$ is the prior of

the parameters under the hypothesis H_i . The marginal likelihood $Pr(Y|H_i)$ can be interpreted as a measure of the data's plausibility under H_i .

Therefore, the Bayes factor, BF_{12} , quantifies the relative support of the data for the two competing hypotheses (and not the ratio between the probability of the two hypotheses). For values close to 1, the Bayes factor indicates that H_1 and H_2 have similar support from the data. Larger values indicate evidence in favor of H_1 whereas values close to 0 indicate evidence in favor of H_2 . The ratio between these two hypotheses' posterior probabilities can be computed as

$$\frac{Pr(H_1|Y)}{Pr(H_2|Y)} = \frac{\overbrace{Pr(Y|H_1)}^{BF_{12}}}{Pr(Y|H_2)} \times \frac{Pr(H_1)}{Pr(H_2)}, \quad (5)$$

where $Pr(H_i)$ is the prior probability of H_i and $Pr(H_i|Y)$ is the posterior probability of H_i . For a detailed description of the Bayes factor, considering also its interpretation and application in various contexts see Heck et al. (2020), Mulder and Wagenmakers (2016), and Wagenmakers et al. (2010).

Note that to compute the marginal likelihood $Pr(Y|H_i)$ of a hypothesis, H_i , it is necessary to integrate the product between the likelihood and the prior. These integrals, however, are usually difficult to compute (in particular when hypotheses include order constraints). In the case of hypotheses with equality and inequality constraints, however, it is possible to simplify the computation of the Bayes factor using the encompassing prior approach.

2.2.1 *Encompassing Prior Approach*

The basic idea of the encompassing prior approach is to consider an informative hypothesis as a subset of an unconstrained model's parameter space. Therefore, to evaluate a hypothesis's plausibility, we can consider the proportion of parameter space of the unconstrained model that satisfies the constraints. More specifically, given an informative hypothesis, H_i , and an unconstrained model, H_u , (or *encompassing model*) that does not contain any constraints on the parameters, if the prior under H_i is defined as a truncation

of the proper prior under H_u (*encompassing prior*) according to the constraints, then the Bayes factor between H_i and H_u can be written as

$$\begin{aligned} BF_{iu} &= \frac{Pr(\text{Inequality Const}|\text{Equality Const, Data, } H_u)}{\pi(\text{Inequality Const}|\text{Equality Const, } H_u)} \times \frac{Pr(\text{Equality Const}|\text{Data, } H_u)}{\pi(\text{Equality Const}|H_u)} \\ &= \frac{Pr(R_I\theta > r_I|R_E\theta = r_E, Y, H_u)}{\pi(R_I\theta > r_I|R_E\theta = r_E, H_u)} \times \frac{Pr(R_E\theta = r_E|Y, H_u)}{\pi(R_E\theta = r_E|H_u)}. \end{aligned} \quad (6)$$

The first term is the ratio between the conditional posterior probability and the conditional prior probability that the inequality constraints hold under the unconstrained model, H_u , given the equality constraints. The second term is the ratio between marginal posterior density and the marginal prior density of the equality constraints under H_u (the well-known Savage–Dickey density ratio; Equation (6) can be seen as an extension of the Savage–Dickey method, where the density ratio of the equality constraints is multiplied by the probability ratio of the posterior and prior. Dickey, 1971; Wetzels et al., 2010). In particular, the four elements can be interpreted as:

- The **conditional posterior probability**, $Pr(R_I\theta > r_I|R_E\theta = r_E, Y, H_u)$, is a measure of the fit of the inequality constraints of H_i under H_u .
- The **conditional prior probability**, $\pi(R_I\theta > r_I|R_E\theta = r_E, H_u)$, is a measure of the complexity of the inequality constraints of H_i under H_u .
- The **marginal posterior density**, $Pr(R_E\theta = r_E|Y, H_u)$, is a measure of the fit of the equality constraints of H_i under H_u .
- The **marginal prior density**, $\pi(R_E\theta = r_E|H_u)$, is a measure of the complexity of the equality constraints of H_i under H_u .

Summarizing at the denominators, we have measures of the **complexity** of the inequality and equality constraints of the informative hypothesis H_i ; at the numerators, instead, we have measures of the **fit** of the data to the inequality and equality constraints

of the informative hypothesis H_i . Where the *fit* of an hypothesis indicates its ability to provide a good description of data at hand and the *complexity* of an hypothesis refers to its level of specificity and parsimony (i.e., simpler hypothesis are specific and parsimonious). Therefore, if the hypothesis H_i , although applying constraints to the parametric space (less complexity), can still provide an accurate description of the data, the Bayes factor will favor H_i . On the contrary, if the support of the data is poor, the Bayes factor will favor the unconstrained model, H_u .

The proof of the formulation of the Bayes factor with the encompassing prior approach and the evaluation of its consistency (i.e., the probability of selecting the correct hypothesis goes to 1 for the sample size going to infinity) is provided in Gu et al. (2018) and Mulder and Gelissen (2019). Note that slightly different notation is used here to enhance comprehension and emphasize that at the numerators, we have values computed from the posterior of H_u whereas at the denominators, we have values computed from the prior of H_u .

To compute the Bayes factor with the encompassing prior approach, only the prior and the posterior of the unconstrained model are required. To obtain them, first, we need to define the encompassing prior of the unconstrained model.

2.2.2 Definition of the Encompassing Prior

Prior specification is an important element as the resulting Bayes factor value is affected by the prior choice. This is particularly relevant in the case of equality constraints (Bartlett, 1957; Lindley, 1957), whereas inequality constraints are not affected as long as the prior is symmetric and centered on the focal point of interest. We will discuss this feature further in Section 2.2.3.

To avoid arbitrary prior specification, researchers have proposed various methods. For example: Jeffreys-Zellner-Siow objective priors do not require subjective specification (Bayarri & García-Donato, 2007; Jeffreys, 1961; Zellner & Siow, 1980); partial Bayes factor (de Santis & Spezzaferri, 1999) defines the prior according to part of the data whereas the

remaining part is used to compute the Bayes factor; intrinsic Bayes factor (Berger & Pericchi, 1996) and fractional Bayes factor (O’Hagan, 1995) are variations on the partial Bayes factor where priors are defined according to the average of all possible minimal subsets of the data or to a given small fraction of the data, respectively.

Gu et al. (2018) and Mulder and Olsson-Collentine (2019) based their approaches on the fractional Bayes factor. Starting from a noninformative prior, a minimal fraction of the data is used to obtain a posterior that is subsequently used as proper prior, which we refer to as *fractional prior*. Consider Lindley’s (1972) famous quote, “*Today’s posterior is tomorrow’s prior*”. The remaining part of the data is used to compute the Bayes factor. These two approaches, however, have an important difference. Gu et al. (2018) approximated the obtained fractional prior (as well as the posterior, see Section 2.2.4) to a (multivariate) normal distribution. In fact, according to large-sample theory (Gelman et al., 2013), parameter posterior distribution can be approximated to a (multivariate) normal distribution. On the contrary, Mulder and Olsson-Collentine (2019) provided an analytic solution in the case of linear regression models, obtaining an accurate quantification of the distribution.

All methods discussed above suggest objective procedures to avoid arbitrary prior specification. Nevertheless, using subjective (reasonable) priors according to previous information or experts’ indications is still a possible approach. However, prior specification (even if obtained through an *objective procedure*) affects the Bayes factor results. Therefore, it is necessary to conduct a prior sensitivity analysis to evaluate the influence of prior specification on the results (Du et al., 2019; Schad, Nicenboim, et al., 2021).

Going back to our psychological treatment example, we could use independent normal distributions to specify the prior of each parameter of interest (i.e. $\theta_{control}$, $\theta_{traditional}$, $\theta_{individual}$, θ_{group} , and θ_{online}), obtaining as resulting prior a multivariate normal distribution with mean vector μ_θ and covariance matrix Σ_θ . In this way, we can still follow the approach Gu et al. (2018) proposed based on normal approximation, which will

simplify the computation of the Bayes factor. For example, suppose that, according to experts' indications, a reasonable prior choice for all parameters of interest is a normal distribution with a mean of 0 and standard deviation of 2: $\mathcal{N}(0, 2)$. The resulting prior is a multivariate normal distribution with mean vector μ_θ and covariance matrix Σ_θ :

$$\pi(\boldsymbol{\theta}) \sim \mathcal{N}(\mu_\theta, \Sigma_\theta)$$

$$\text{where } \boldsymbol{\theta} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \Sigma_\theta = \begin{bmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}. \quad (7)$$

Note that, prior specification is relevant only for the parameters of interest (i.e., the parameters involved in the constraints). Nuisance parameters (i.e., the parameters not involved in the constraints) are integrated out in the Bayes factor. Thus, the results will be insensitive to the actual choice of the prior for the nuisance parameters as long as it is vague enough.

At this point, several authors emphasize the importance of centering the prior distribution on the constraints' points of interest. Let's further discuss this issue in the next section.

2.2.3 *Adjusting Prior Mean*

When researchers evaluates informative hypotheses with equality and inequality constraints using the Bayes factor, they must center the priors on the focal points of interest (Jeffreys, 1961; Mulder, 2014; Zellner & Siow, 1980). In the case of equality constraints, centering the prior allows one to consider values close to the point of interest more likely a priori than distant values. This adjustment should be in line with researchers' expectations; otherwise, one could question why he or she is testing that value.

In the case of inequality constraints, instead, this adjustment is done to guarantee that no constraint is favored a priori. Consider the example represented in Figure 1, where the hypothesis $\theta_i > k$ is evaluated against $\theta_i < k$. Remember that, when computing the Bayes factor, the prior probability that the constraint holds is used as a measure of the hypothesis's complexity. Therefore, if the prior is not centered on the focal point of interest (i.e., k), the less complex hypothesis (i.e., $\theta_i > k$) will be erroneously preferred a priori over the other (i.e., $\theta_i < k$). The hypotheses will be equally likely a priori only when the prior is centered on the focal point.

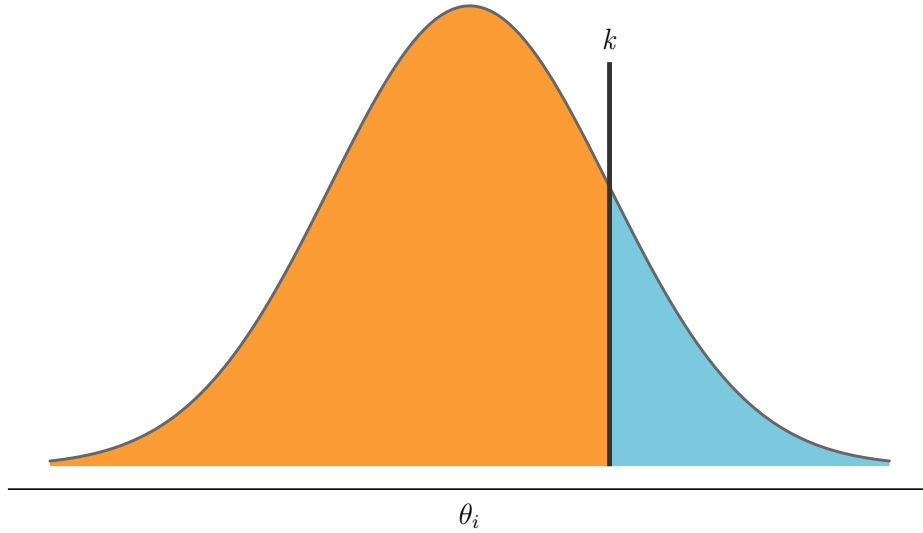


Figure 1

Example of non centered prior considering the constraints $\theta_i > k$ vs. $\theta_i < k$.

Centering the prior to the focal point, however, can be difficult in the case of complex hypotheses in which constraints are defined as a linear combination of several parameters. To overcome this issue, Gu et al. (2018) proposed the following transformation

of the parameters of interest:

$$\boldsymbol{\beta} = R\boldsymbol{\theta} - r$$

$$\text{with } \boldsymbol{\beta} = \begin{bmatrix} \beta_E \\ \beta_I \end{bmatrix}, \quad R = \begin{bmatrix} R_E \\ R_I \end{bmatrix} \quad \text{and} \quad r = \begin{bmatrix} r_E \\ r_I \end{bmatrix}, \quad (8)$$

where $\boldsymbol{\theta}$ is the vector of original parameters, R is the matrix expressing equality and inequality constraints, and r is the vector with the constants of the equality and inequality constraints. The informative hypothesis under evaluation then becomes:

$$H_i : \quad \beta_E = 0 \quad \& \quad \beta_I > 0. \quad (9)$$

This parameter transformation has the advantage of simplifying the hypothesis expression without changing the original expectations. In fact, for example, evaluating $\theta_i > \theta_j$ is equivalent to evaluating $\beta_i = \theta_i - \theta_j > 0$. Therefore, given the original prior $\pi(\boldsymbol{\theta}) \sim \mathcal{N}(\mu_\theta, \Sigma_\theta)$, the prior of the new parameter vector $\boldsymbol{\beta}$ is given by

$$\pi(\boldsymbol{\beta}) \sim \mathcal{N}(\mu_\beta, \Sigma_\beta) = \mathcal{N}(R\boldsymbol{\theta} - r, R\Sigma_\theta R^T). \quad (10)$$

Note that this operation is nothing more than applying a linear transformation to the original multivariate normal distribution.¹ To do that, however, the matrix R must be *full-row-rank* (i.e., all rows are linearly independent). If this is not the case, the obtained matrix Σ_β will not be a proper covariance matrix. In Section 2.4.3, we will discuss a solution to overcome this limit.

Now, to center the prior of $\boldsymbol{\beta}$ to the focal points, we can simply set the mean vector to 0. Therefore, the adjusted prior of $\boldsymbol{\beta}$ is

$$\pi_{adj}(\boldsymbol{\beta}) \sim \mathcal{N}(\mathbf{0}, \Sigma_\beta) = \mathcal{N}(\mathbf{0}, R\Sigma_\theta R^T). \quad (11)$$

¹ Given a multivariate normal distribution $Y \sim \mathcal{N}(\mu, \Sigma)$, the result of a linear transformation $AY + b$ is still a multivariate normal distribution with vector mean $\mu_t = AY + b$ and covariance matrix $\Sigma_t = A\Sigma A^T$.

Considering the psychological treatment example, the obtained adjusted prior is $\mathcal{N}(\mathbf{0}, \Sigma_\beta)$, where:

$$\mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \Sigma_\beta = R\Sigma_\theta R^T = \begin{bmatrix} 8 & -4 & 0 & 4 \\ -4 & 8 & 0 & 0 \\ 0 & 0 & 8 & -4 \\ 4 & 0 & -4 & 8 \end{bmatrix}. \quad (12)$$

So far, we have defined the prior for the parameter vector $\boldsymbol{\theta}$ of the encompassing model. Moreover, we have obtained the adjusted prior for the new transformed parameter vector $\boldsymbol{\beta}$, allowing us to properly evaluate the equality and inequality constraints. At this point, we need to compute the posterior of the encompassing model parameters.

2.2.4 *Posterior Encompassing Model*

Posterior distribution of the encompassing model parameters can be obtained through numerical approximation using Markov Chain Monte Carlo (MCMC) sampling algorithms, such as the Metropolis–Hastings algorithm (Hastings, 1970) or Gibbs sampling (Geman & Geman, 1984). Bayesian statistical inference methods are implemented in all major statistical software. In R statistical software (R Core Team, 2021), for example, the popular `brms` package (Bürkner, 2017, 2018), which is based on STAN (Stan Development Team, 2020), allows researchers to easily conduct Bayesian inference.

Following the approach established by Gu et al. (2018), once we obtain the model posterior, we can approximate it to a (multivariate) normal distribution. Therefore, the resulting posterior distribution is

$$Pr(\boldsymbol{\theta}|Y) \sim \mathcal{N}(\hat{\boldsymbol{\theta}}, \hat{\Sigma}_\theta), \quad (13)$$

where posterior mean $\hat{\boldsymbol{\theta}}$ and posterior covariance $\hat{\Sigma}_\theta$ can be computed directly from the posterior draws. Next, we can obtain the posterior with respect to the vector parameters $\boldsymbol{\beta}$ applying the same transformation used for the prior distribution:

$$Pr(\boldsymbol{\beta}|Y) \sim \mathcal{N}(\hat{\boldsymbol{\beta}}, \hat{\Sigma}_\beta) = \mathcal{N}(R\hat{\boldsymbol{\theta}} - \mathbf{r}, R\hat{\Sigma}_\theta R^T). \quad (14)$$

At this point, we have both prior and posterior distributions of the parameters of interests of the encompassing model. Before proceeding, however, let's emphasize some important aspects.

When defining the prior, we adjusted the prior mean by centering it over the constraint's focal points. This change would slightly influence the resulting posterior, as well. However, as Gu et al. (2018) pointed out, small prior changes will result in negligible changes in the posterior for large samples, according to large-sample theory. Therefore, the authors do not adjust the posterior in their approach but only the prior. Therefore, when computing the posterior, we can simply consider the prior $\pi(\boldsymbol{\theta}) \sim \mathcal{N}(\mu_{\theta}, \Sigma_{\theta})$ defined at the beginning without worrying about adjusting it.

In addition, note that in the original approach Gu et al. (2018) proposed, the posterior mean $\hat{\boldsymbol{\theta}}$ and the posterior covariance $\hat{\Sigma}_{\theta}$ are not obtained from the posterior draws but are computed directly from the sample data using the maximum likelihood estimate and the inverse of the Fisher information matrix, respectively. This method has the advantage of being faster (posterior draws are not required), but it may be not possible for some complex models.

2.3 Computing the Bayes Factor

In the previous section, we obtained the adjusted prior and the posterior of the vector of transformed parameters $\boldsymbol{\beta}$, respectively

$$\pi_{adj}(\boldsymbol{\beta}) \sim \mathcal{N}(0, \Sigma_{\beta}) \quad \text{and} \quad Pr(\boldsymbol{\beta}|Y) \sim \mathcal{N}(\hat{\boldsymbol{\beta}}, \hat{\Sigma}_{\beta}). \quad (15)$$

Given the parameter transformation from $\boldsymbol{\theta}$ to $\boldsymbol{\beta}$, we can rewrite Formula 6 of the Bayes factor between H_i and H_u as follows:

$$BF_{iu} = \frac{Pr(\beta_I > 0 | \beta_E = 0, Y, H_u)}{\pi_{adj}(\beta_I > 0 | \beta_E = 0, H_u)} \times \frac{Pr(\beta_E = 0 | Y, H_u)}{\pi_{adj}(\beta_E = 0 | H_u)}, \quad (16)$$

where β_I and β_E are defined in Formula 8 and represent the inequality and equality constraints, respectively.

Now, thanks to (multivariate) normal distribution approximation, we can easily compute the required conditional probabilities and marginal densities required to calculate the Bayes factor.

2.3.1 Marginal Density

The marginal distribution of a subset of variables of a multivariate normal distribution is obtained by simply discarding the variables to marginalize out. For example, given the adjusted prior of the psychological treatment example, $\pi_{adj}(\boldsymbol{\beta}) \sim \mathcal{N}(\mathbf{0}, \Sigma_{\beta})$, where

$$\mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \Sigma_{\beta} = \begin{bmatrix} 8 & -4 & 0 & 4 \\ -4 & 8 & 0 & 0 \\ 0 & 0 & 8 & -4 \\ 4 & 0 & -4 & 8 \end{bmatrix}, \quad (17)$$

the marginal distribution of the equality constraints $\boldsymbol{\beta}_E$ is

$$\pi_{adj}(\boldsymbol{\beta}_E) \sim \mathcal{N}(\mu_{\beta_E}, \Sigma_{\beta_E}), \quad (18)$$

$$\text{where } \mu_{\beta_E} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \Sigma_{\beta_E} = \begin{bmatrix} 8 & -4 \\ -4 & 8 \end{bmatrix}.$$

At this point, computing the density at $\boldsymbol{\beta}_E = \mathbf{0}$ is elementary. In R, this can be done using the `dmvnorm()` function from the `mvtnorm` package (Genz et al., 2021). To compute the **marginal prior density**, $\pi_{adj}(\boldsymbol{\beta}_E = \mathbf{0} | H_u)$, of the psychological treatment example, we can use the following code:

```
# Prior info
mu_prior <- c(0, 0, 0, 0)
Sigma_prior <- matrix(c( 8,-4, 0, 4,
                        -4, 8, 0, 0,
                        0, 0, 8,-4,
                        4, 0,-4, 8), ncol = 4, byrow = TRUE)

# Marginal prior density at beta_1 = 0 and beta_2 = 0
```

```

mvtnorm::dmvnorm(x = c(0, 0),
                 mean = mu_prior[1:2],
                 sigma = Sigma_prior[1:2, 1:2])

## [1] 0.02297204

```

Similarly, it is possible to compute the **marginal posterior density**, $\pi_{adj}(\beta_E = 0|Y, H_u)$, considering this time the estimated posterior mean vector, $\hat{\beta}$, and the estimated posterior covariance matrix, $\hat{\Sigma}_\beta$.

2.3.2 Conditional Probability

To compute the conditional probability of the inequality constraints given the equality constraints in a multivariate normal distribution, we can use the `pcmvnorm()` function from the `condMVNorm` R-package (Varadhan, 2020). Considering the psychological treatment example, to calculate the **conditional prior probability**,

$\pi_{adj}(\beta_I > 0|\beta_E = 0, H_u)$, we can run the following code:

```

# Conditional prior probability that beta_3 > 0 and beta_4 > 0
# given beta_1 = 0 and beta_2 = 0
condMVNorm::pcmvnorm(
  lower = c(0, 0), upper = c(Inf, Inf), # inequality constraints
  mean = mu_prior, sigma = Sigma_prior,
  dependent.ind = 3:4,                  # inequality variables
  given.ind = 1:2, X.given = c(0, 0))  # equality variables and constraints

## [1] 0.1451077
## attr(,"error")
## [1] 1e-15
## attr(,"msg")
## [1] "Normal Completion"

```

Analogously, it is possible to compute the **conditional posterior probability**, $\pi_{adj}(\beta_I > 0|\beta_E = 0, Y, H_u)$, considering this time the estimated posterior mean vector, $\hat{\beta}$, and the estimated posterior covariance matrix, $\hat{\Sigma}_\beta$.

At this point, we have all the elements required, and we can easily compute the Bayes factor, BF_{iu} .

2.4 Advanced Elements

In the previous sections, we have described in detail all the steps and elements required to evaluate informative hypotheses using the Bayes factor with the encompassing prior approach, following the approximated method Gu et al. (2018) proposed. As already highlighted, however, compared to the original approach Gu et al. (2018) proposed, we took a slightly different path in the prior definition and posterior approximation. Instead of using the fractional prior method, we defined subjective (reasonable) priors, and we relied on posterior draws instead of maximum likelihood estimates to approximate the posterior. We made these choices to follow the usual Bayesian workflow in data analysis (Gelman et al., 2020; Schad, Betancourt, et al., 2021).

Before moving to an applied example, however, we consider some more advanced aspects that are important to take into account when applying this method. In particular, we discuss issues related to parameters' standardization, range constraints, hypothesis matrix rank, comparable hypotheses, and bounded parameter space.

2.4.1 *Parameters Standardization*

When evaluating informative hypotheses, it may be necessary to standardize the parameters of interest. As Gu et al. (2018) showed, parameters have to be standardized when comparing regression coefficients of continuous variables that are measured on different scales - in these cases, in fact, determining whether $\theta_i > \theta_j$ is affected by the scale of the two variables. Standardizing the parameters allows us to correctly compare the two parameters. On the contrary, standardizing the parameters is not necessary if they are compared to constants (i.e., $\theta > 3$) or if they represent group means, such as in the case of categorical variables.

2.4.2 *Range Constraints*

It is possible to express range constraints, such as $0 < \theta_i < 1$, in which the parameter is constrained between two values. These constraints can be expressed as the combination of two constraints. For example, the previous hypothesis is equivalent to $\beta_1 = \theta_i > 0$ and

$\beta_2 = 1 - \theta_i > 0$. However, adjusting the prior mean in the case of range constraints would require us to center the prior on the middle of the range space (Mulder et al., 2012).

Therefore, in the previous example, we would set $\mu_{\beta_1} = \mu_{\beta_2} = .5$. For a detailed description of prior specification in the case of range constraints, see Appendix A in Gu et al. (2018).

2.4.3 Hypothesis Matrix Rank

In Equation 10, we presented how to obtain the distribution of the transformed parameters β from the distribution of the original parameters θ . This is simply a linear transformation, but to obtain a proper covariance matrix $\Sigma_\beta = R\Sigma_\theta R^T$, the hypothesis matrix R has to be *full-row-rank* (i.e., all rows linearly independent).

To overcome this limitation, we can use the same solution adopted by Mulder and Olsson-Collentine (2019) and Mulder (2016). A matrix R^* is defined by selecting the maximum number of linearly independent rows of R and used to compute the covariance matrix $\Sigma_\beta = R^*\Sigma_\theta R^{*T}$. This allows us to obtain the distribution of the parameters β^* that are linearly independent. The remaining constraints can be expressed as linear combinations of the parameters β^* . Finally, to evaluate the probability that all constraints hold, we can draw a large sample from the distribution of β^* and compute the proportion of draws satisfying all the constraints.

The following is an example to clarify this procedure. Suppose the following hypothesis: $H_i : 0 = \theta_1 < \{\theta_2, \theta_3\} < \theta_4$. Thus, the corresponding hypothesis matrix is

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \quad \text{with} \quad \begin{aligned} \beta_1 &= \theta_1 \\ \beta_2 &= \theta_2 - \theta_1 \\ \beta_3 &= \theta_3 - \theta_1 \\ \beta_4 &= \theta_4 - \theta_2 \\ \beta_5 &= \theta_4 - \theta_3 \end{aligned} \quad (19)$$

Note that only the first four rows ($\beta_1, \beta_2, \beta_3, \beta_4$) are linearly independent. The fifth row can be obtained as a linear combination: $\beta_5 = \beta_4 - \beta_3 + \beta_2$. At this point, we can

define R^* as the first four rows of R and use it to obtain the distribution of β^* given the original distribution of θ . To compute the probability of the inequality constraints, we need to draw a large sample from the multivariate normal distribution conditional to the equality constraints. To do that, we can use the `rcmvnorm()` function from the R-package `condMVNorm`.

```
# theta distribution
theta_mu <- c(0,0,0,0)
theta_sigma <- diag(4)*2

# Hypothesis matrix
R <- matrix(c(1, 0, 0, 0,
              -1, 1, 0, 0,
              -1, 0, 1, 0,
              0,-1, 0, 1,
              0, 0,-1, 1), ncol = 4, byrow = TRUE)
r <- c(0, 0, 0, 0, 0)

# beta distribution
beta_mu <- R[1:4,] %*% theta_mu - r[1:4]
beta_sigma <- R[1:4, ] %*% theta_sigma %*% t(R[1:4, ])

# Obtain draws
set.seed(2021)
obs <- condMVNorm::rcmvnorm(
  1e4, mean = beta_mu, sigma = beta_sigma,
  dependent.ind = 2:4,          # inequality variables
  given.ind = 1, X.given = 0) # equality variable and constraint

# Define linearly dependent constraints
colnames(obs) <- c("beta_2", "beta_3", "beta_4")
obs <- cbind(obs,
             "beta_5" = obs[, "beta_4"] - obs[, "beta_3"] + obs[, "beta_2"])

# Evaluate on each draw if all constraints hold
test <- rowSums(obs > 0) == 4

# Proportion
mean(test)

## [1] 0.0428
```

2.4.4 *Comparable Hypotheses*

To compute the Bayes factor between two different informative hypotheses, we can simply consider the ratio of the Bayes factor of each hypothesis against the encompassing model:

$$BF_{ij} = \frac{BF_{iu}}{BF_{ju}} \quad (20)$$

To do that, however, it is necessary for the informative hypotheses considered to be comparable (Gu et al., 2018). That is, the intersection of the constrained region of each hypothesis must be nonempty. Thus, a set of informative hypotheses $\{H_1, \dots, H_i\}$ is comparable if exist at least one solution to

$$\begin{bmatrix} R_{1E} \\ R_{1I} \end{bmatrix} \theta = \begin{bmatrix} r_{1E} \\ r_{1I} \end{bmatrix}, \dots, \begin{bmatrix} R_{iE} \\ R_{iI} \end{bmatrix} \theta = \begin{bmatrix} r_{iE} \\ r_{iI} \end{bmatrix}. \quad (21)$$

This will allow us to select a common solution of θ that can be used as the adjusted prior mean of the encompassing model. Note that to solve these equations, inequality constraints are also considered equality constraints.

For example, $H_1 : \theta_1 > \theta_2$ and $H_2 : 0 = \theta_1 < \theta_2$ are comparable hypotheses because a common solution exists ($\theta_1 = \theta_2 = 0$). On the contrary, $H_3 : \theta_1 < \theta_2 + 3$ and $H_4 : \theta_1 = \theta_2$ are not comparable hypotheses because there is no common solution to $\theta_1 = \theta_2 + 3$ and $\theta_1 = \theta_2$.

2.4.5 *Bounded Parameter Space*

The described approach using (multivariate-) normal approximation can be straightforwardly applied when parameters of interest are unbounded. However, in the case of bounded parameters (i.e., variances, correlations, or probabilities), appropriate prior distributions are required to take into account the parameters' bounded nature.

Note that if only inequality constraints are considered, the prior distribution is only used to measure the complexity of the inequality-constrained hypotheses. In this case, the prior probability that an inequality constraint holds will be insensitive to the exact

distributional form, as long as the distribution is symmetrical and centered on the focal point. For example, the probability that a correlation coefficient ρ_i is greater than zero is the same if we consider a uniform distribution between -1 and 1 (taking into account parameter's bounded nature) or if we consider a normal distribution centered on zero (without taking into account parameter's bounded nature; see Figure 2).

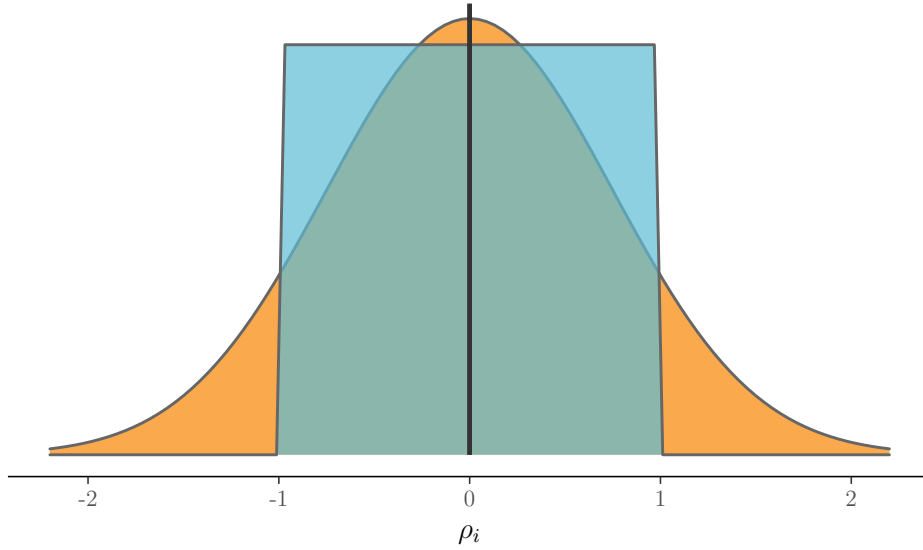


Figure 2

The probability that $\rho_i > 0$ is the same under the uniform prior and the normal prior.

Thus, as suggested by Gu et al. (2014), as long as the complexity of the constrained hypotheses is the same regardless of using (multivariate-) normal distribution or an appropriate bounded prior distribution, it is still possible to follow the described approach. This is true for hypotheses that belong to an equivalent set (for a rigorous definition of equivalent sets, see Gu et al., 2014; Hoijtink, 2012).

In the case of equality and inequality constraints, however, normal approximation is possible. This is done based on the same rationale as before and is only used to compute the conditional prior probability. To estimate the marginal prior density, alternative solutions have to be adopted using appropriate prior distribution to take the bounded nature of the parameters into account. See Mulder and Gelissen (2019) for an example of

hypotheses evaluating correlation coefficients in generalized multivariate probit models.

These issues are relevant only when considering the prior distribution. In the case of the posterior distribution, instead, we can safely approximate it to a (multivariate) normal distribution thanks to large-sample theory. Moreover, prior specification is relevant only for the parameters of interest (i.e., the parameters involved in the constraints), as nuisance parameters are integrated out in the Bayes factor. Thus, the results will be insensitive to the actual choice of the prior for the nuisance parameters as long as it is vague enough.

3 A Case Study: Hypotheses Testing in Attachment Theory

In this section, we propose a real case study evaluating different informative hypotheses within attachment theory. This will allow us to face the common issues found in real research applications. First, we provide the required background information regarding attachment theory and the study aims and characteristics. Subsequently, we describe all the steps involved in the evaluation of informative hypotheses using the Bayes factor with the encompassing prior approach. Finally, we briefly discuss the obtained results, considering the prior sensitivity analysis and limitations of this approach.

All analyses were conducted using R statistical software (Version 4.2.1; R Core Team, 2021). All materials, data, and analysis code are available at <https://doi.org/10.17605/OSF.IO/JKMDF>, GitHub repository <https://github.com/ClaudioZandonella/Attachment>. Supplemental Material with further details is also available online via GitHub Pages at <https://claudiozandonella.github.io/Attachment/>.

Note that data used for the illustrative application are part of a larger project on the social-emotional development of school-age children, which was approved by the Ethics Committee of the School of Psychology at the University of Padova (protocol #[1838-2016]).

3.1 Background Information

3.1.1 *Attachment Theory*

Attachment theory originates from the pioneering work of Bowlby (1969) and Ainsworth (1970). They postulated that children in stressful situations actively seek proximity of the caregiver through attachment behaviors (e.g., crying; moving toward the caregiver) in order to fulfil the evolutionary goal of protection from dangers. The main tenet of attachment theory is that the relationships with the caregivers that children develop in the early stages of their lives (i.e., *attachment bonds*) will affect children's social-emotional future development (Cassidy & Shaver, 2016). Besides behavior, people construct mental representations, or working models, of the self and significant others based on interpersonal experiences. Describing individual differences in self-reported attachment mental representations, empirical research suggested a model based on two dimensional traits (Mikulincer & Shaver, 2007): attachment avoidance (characterized by discomfort with closeness, excessive self-reliance, and difficulty in relying on others for obtaining security) and attachment anxiety (characterized by low self-worth and fear of abandonment and rejection). This perspective, initially developed with adult participants, has been extended to the study of self-reported attachment representations during childhood as well as during middle childhood (Kerns & Richardson, 2005; Marci et al., 2019). The two dimensions of attachment mental representations can be combined to provide four different attachment styles:

- **Secure Attachment:** Children who are securely attached display optimal emotional regulation and consider their caregiver a secure base.
- **Anxious Attachment:** Anxious children manifest high levels of anxiety in stressful situations; their relationships with caregivers are ambivalent, displaying anger or helplessness.

- **Avoidant Attachment:** Avoidant children mask distress in stressful situations, displaying few emotions; their relationships with caregivers are characterized by little involvement.
- **Fearful Attachment:** Fearful children lack adequate emotional regulation in stressful situations, with the risk of displaying disorganized behaviors.

Attachment theory is one of the most supported theories in psychology (Cassidy & Shaver, 2016). In the literature, however, there is still an open debate on the relative influence of mother and father attachment on children’s social-emotional development. Four different main theoretical perspectives have been identified (Bretherton, 2010):

- **Monotropy Theory:** Only the principal attachment figure (usually the mother) has an impact on children’s development.
- **Hierarchy Theory:** The principal attachment figure (usually the mother) has a greater impact on children’s development than the subsidiary attachment figure (usually the father).
- **Independence Theory:** All attachment figures are equally important, but they affect the children’s development differently.
- **Integration Theory:** To understand their impact on children’s development, it is necessary to consider all attachment relationships together.

Contrasting results have been reported by studies investigating which is the “*correct*” theory. No study, however, has tried to properly evaluate the different theoretical perspectives by directly comparing the different hypotheses.

3.1.2 Present Study

We aimed to directly compare the four different theoretical perspectives regarding the influence of father and mother attachment using the Bayes factor with the encompassing prior approach.

Out of our initial population, 847 Italian children (50.65% females) between 8 and 12 years old (*middle childhood*, third to sixth grade) were included in the final analysis. Attachment toward the mother and the father were measured separately using the brief Experiences in Close Relationships Scale - Revised Child Version (ECR-RC; Brenning et al., 2014; Marci et al., 2019) completed by the children. Subsequently, we performed two separate cluster analyses for mother and father scores. Both analyses supported the existence of the four attachment profiles (see above). The results of the classification are reported in Table 1.

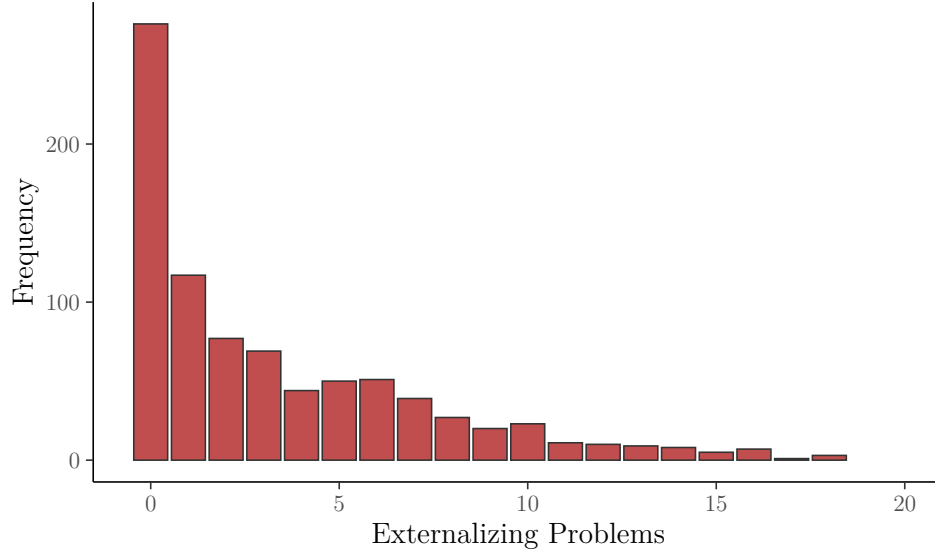
Table 1

Attachment styles frequencies ($n_{subj} = 847$).

Mother Attachment	Father Attachment				Total
	Secure	Anxious	Avoidant	Fearful	
Secure	125	49	49	8	231
Anxious	51	100	98	37	286
Avoidant	25	67	126	12	230
Fearful	5	14	38	43	100
Total	206	230	311	100	847

We measured children's social-emotional development using the Strengths and Difficulties Questionnaire (SDQ; Goodman et al., 2010) completed by teachers. Separate scores for externalizing and internalizing problems were obtained as sums of the questionnaire items. For the purpose of this work, we considered only externalizing problems. The distribution of externalizing problems ($M = 3.35$; $SD = 3.91$; Median = 2.0) is presented in Figure 3.

Externalizing problems according to attachment style are reported in Table 2. More information about the sample, descriptive statistics, cluster analysis, and analysis of internalizing problems can be found in the Supplemental Material available online <https://claudiozandonella.github.io/Attachment/>. All materials and data are available at <https://doi.org/10.17605/OSF.IO/JKMDF>, GitHub repository <https://github.com/ClaudioZandonella/Attachment>.

**Figure 3**

Distribution of externalizing problems ($n_{subj} = 847$).

Table 2

Externalizing problems according to attachment styles ($n_{subj} = 847$).

Mother Attachment	Father Attachment							
	Secure		Anxious		Avoidant		Fearful	
	Mean (SD)	Median	Mean (SD)	Median	Mean (SD)	Median	Mean (SD)	Median
Secure	2.63 (3.57)	1.0	3.45 (4.48)	2.0	1.61 (2.13)	1.0	2.88 (3.44)	2.0
Anxious	3.69 (4.07)	2.0	3.01 (3.61)	2.0	3.32 (4.12)	2.0	4.05 (3.61)	3.0
Avoidant	2.84 (3.34)	1.0	3.31 (3.65)	2.0	3.71 (4.19)	2.0	3.75 (4.81)	1.0
Fearful	7.60 (4.04)	8.0	4.64 (3.84)	4.5	4.76 (4.67)	3.0	4.26 (4.07)	4.0

3.2 Evaluating Hypotheses with Bayes Factor

3.2.1 Formalization of Informative Hypotheses

Each theoretical perspective was formalized into a different informative hypothesis taking into account its own theoretical tenets, main evidence in the literature, and our clinical experience in the field.

We used the following notation to formalize the hypotheses. M and F are used to indicate attachment toward the mother and the father, respectively. Specific attachment styles are specified using subscript: S indicates secure attachment, Ax anxious attachment, Av avoidant attachment, and F fearful attachment. For example, F_{Av} represents children

with avoidant attachment toward their fathers. Moreover, * subscript is used to indicate any attachment style and a set of subscripts is used to indicate “*one among*”. For example, F_* indicates children with any type of attachment toward their fathers and $F_{Ax;Av}$ represents children with anxious or avoidant attachment toward their fathers.

Note that when we do not expect interaction between mother and father attachment, we can consider the roles of the two parents separately. Whereas, if an interaction is expected, it is necessary to take into account the unique combination of mother and father attachment. Thus, for example, we use M_SF_{Ax} to indicate children with secure attachment toward their mothers and anxious attachment toward their fathers.

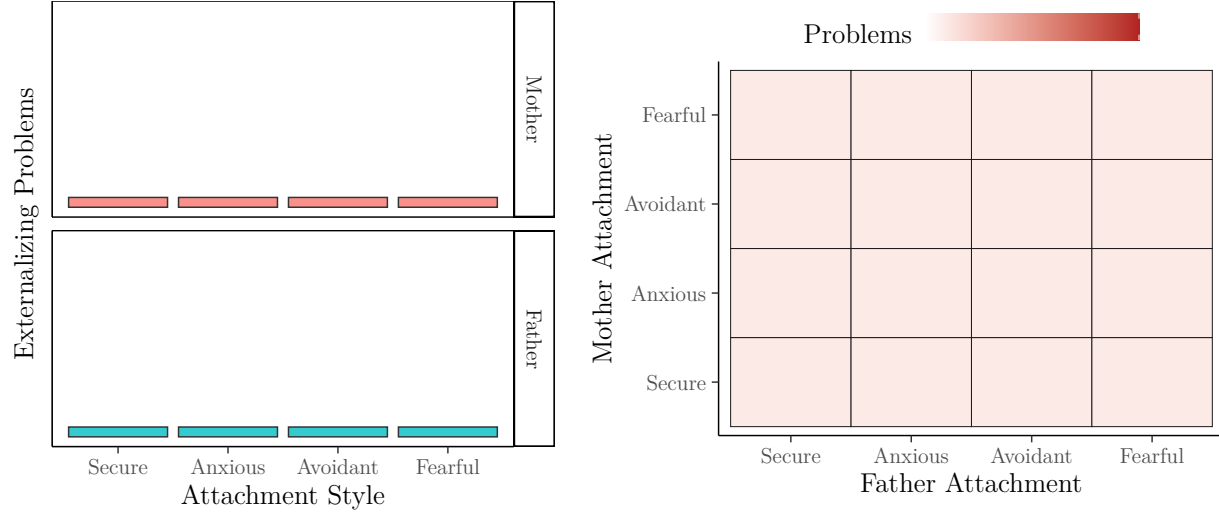
To correctly interpret the following Figures, note that the order is important, whereas the actual values are only indicative.

Null Hypothesis. This is a reference hypothesis in which mother attachment and father attachment are expected to have no effect and only gender differences are taken into account (see Section 3.2.2 for the actual model definition). The hypothesis is represented in Figure 4:

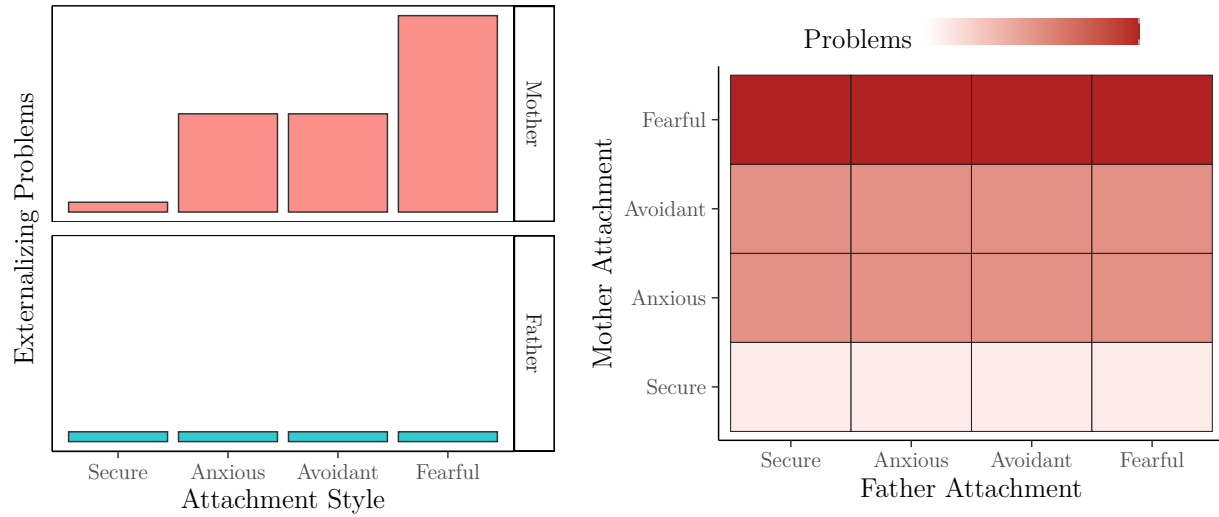
$$\begin{aligned} M_* &= 0, \\ F_* &= 0. \end{aligned} \tag{22}$$

Monotropy Hypothesis. Father attachment is expected to have no effect, whereas considering mother attachment, we expect the following order: secure children should have the lowest level of problems, anxious and avoidant children should exhibit similar levels of problems, and fearful children should exhibit the highest levels of problems. This hypothesis is represented in Figure 5:

$$\begin{aligned} M_S &< M_{Ax} = M_{Av} < M_F, \\ F_* &= 0. \end{aligned} \tag{23}$$

**Figure 4**

Null Hypothesis. Expected externalizing problems according to mother and father attachment.

**Figure 5**

Monotropy Hypothesis. Expected externalizing problems according to mother and father attachment.

Hierarchy Hypothesis. Father attachment is expected to follow the same pattern as mother attachment, but with lesser influence. This hypothesis is represented in Figure 6:

$$M_S < M_{Ax} = M_{Av} < M_F,$$

$$F_S < F_{Ax} = F_{Av} < F_F, \quad (24)$$

$$F_{Ax} < M_{Ax}; \quad F_{Av} < M_{Av}; \quad F_F < M_F.$$

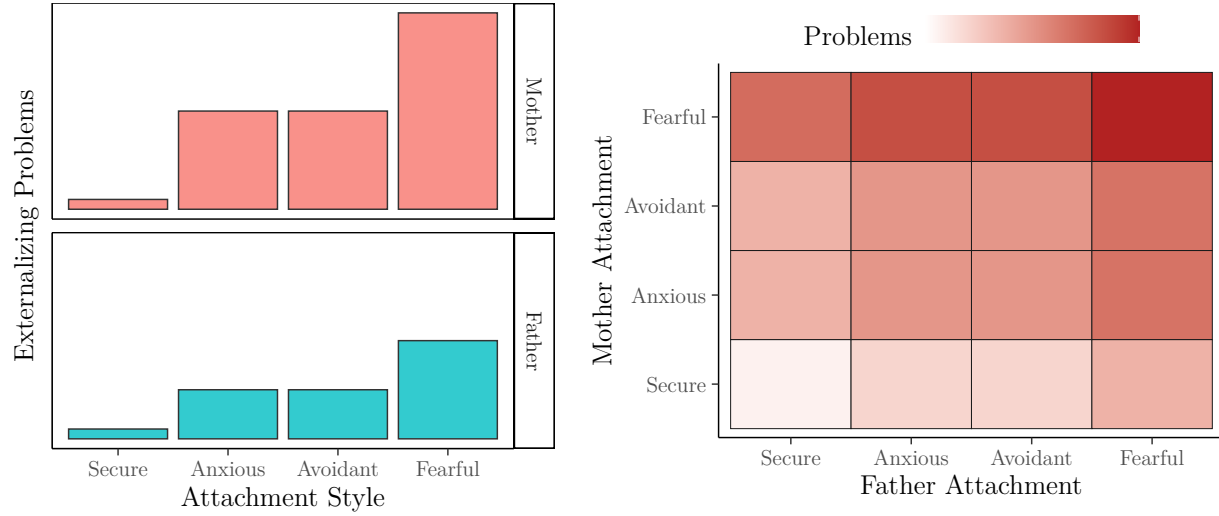


Figure 6

Hierarchy Hypothesis. *Expected externalizing problems according to mother and father attachment.*

Independence Hypothesis. Mother and father attachment are expected to affect children's outcomes differently. In this case, we consider avoidant attachment toward the father as a condition of greater risk. This hypothesis is represented in Figure 7:

$$\begin{aligned}
 M_S &< M_{Ax} = M_{Av} < M_F, \\
 F_S &< F_{Ax} < F_{Av} < F_F.
 \end{aligned}
 \tag{25}$$

Integration Hypothesis. Mother and father attachment are expected to interact. In this case, we consider secure attachment as a protective factor and fearful attachment as a risk condition. We do not specify the conditions M_SF_F and M_FF_S , as their frequency is very low (1.5% of the sample). The hypothesis is represented in Figure 8:

$$M_SF_S < \{M_SF_{Ax;Av} = M_{Ax;Av}F_S\} < M_{Ax;Av}F_{Ax;Av} < \{M_FF_{Ax;Av} = M_{Ax;Av}F_F\} < M_FF_F,$$

with

$$M_{Ax}F_{Ax} = M_{Ax}F_{Av} = M_{Av}F_{Ax} = M_{Av}F_{Av}. \tag{26}$$

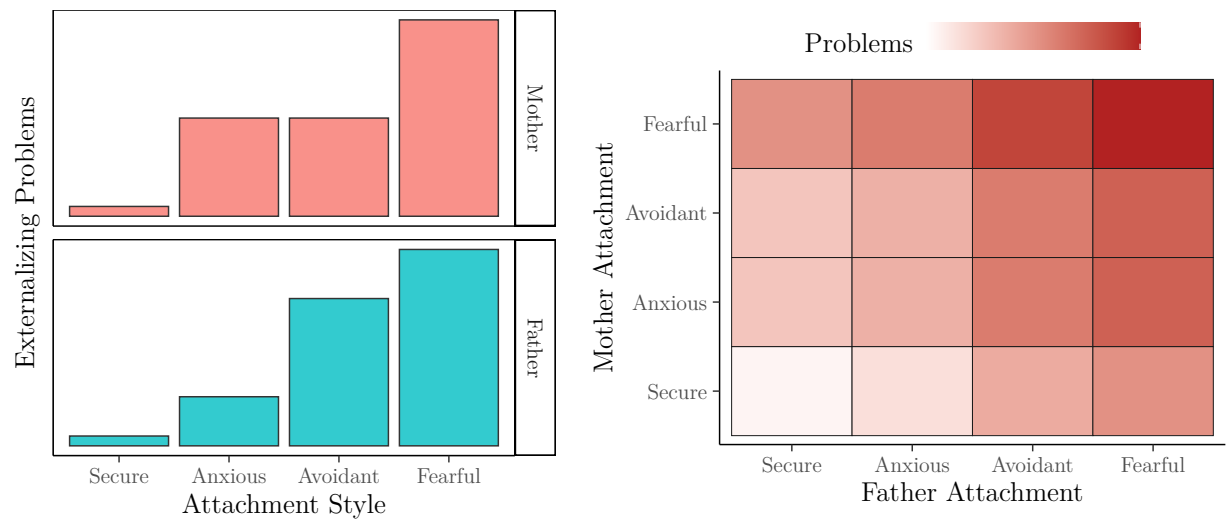


Figure 7

Independence Hypothesis. Expected externalizing problems according to mother and father attachment.

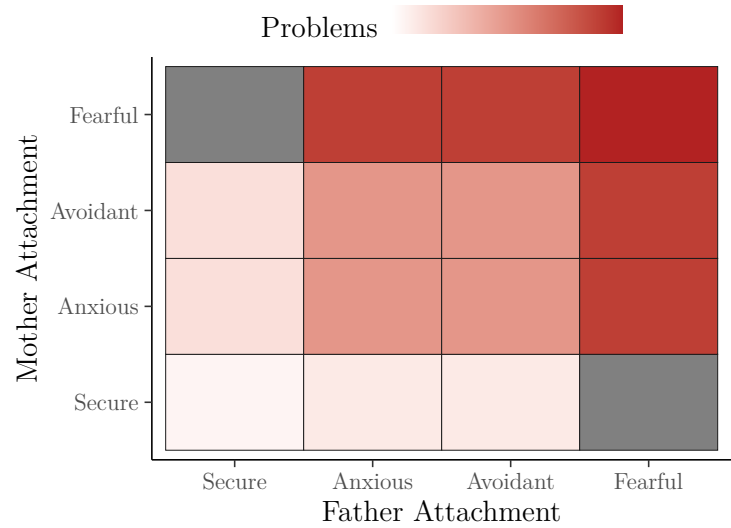


Figure 8

Integration Hypothesis. Expected externalizing problems according to mother and father attachment.

3.2.2 Definition of the Encompassing Model

We define a zero-inflated negative binomial (ZINB) mixed-effect model to take into account the characteristics of the dependent variable and its distribution (see Figure 3 and see Supplemental Material for more details regarding the zero-inflation analysis <https://claudiozandonella.github.io/Attachment/>):

$$y_{ij} \sim ZINB(p_{ij}, \mu_{ij}, \phi), \quad (27)$$

where p_{ij} is the probability of an observation y_{ij} being an extra zero (i.e., a zero not coming from the negative binomial distribution) and $1 - p_{ij}$ indicates the probability of a given observation y_{ij} being generated from a negative binomial distribution with mean μ_{ij} and variance $\sigma_{ij}^2 = \mu_{ij} + \frac{\mu_{ij}^2}{\phi}$. Moreover, we define

$$\begin{aligned} p_{ij} &= \text{logit}^{-1}(X_i^T \beta_p + Z_j^T u_p), \\ \mu_{ij} &= \exp(X_i^T \beta_\mu + Z_j^T u_\mu). \end{aligned} \quad (28)$$

That is, both \mathbf{p} and $\mathbf{\mu}$ are modeled separately according to fixed and random effects. In particular, we consider the children's classroom IDs as a random effect in both cases to account for teachers' different ability to evaluate children's problems. Regarding fixed effects, only the role of gender is considered for \mathbf{p} ; for $\mathbf{\mu}$, the interaction between mother and father attachment are included together with gender. In R formula syntax, we have

```
# Regression on p
p ~ gender + (1|ID_class)

# Regression on mu
mu ~ gender + mother * father + (1|ID_class)
```

The parameters of interest (i.e., those related to mother and father attachment interaction) and the parameter related to gender differences are unbounded. Thus, we can simply specify a normal distribution with a mean of 0 and standard deviation of 3, $\mathcal{N}(0, 3)$, as reasonable prior. This prior is intended to be noninformative but without being

excessively diffuse.² We subsequently evaluated the influence of prior specification in a prior sensitivity analysis. Regarding the other nuisance parameters (i.e., intercepts, random effects, and shape parameters) **brms**' default priors are maintained (see Supplemental Material for further details <https://claudiozandonella.github.io/Attachment/>). The encompassing model was estimated using six independent chains with 10,000 iterations (warm-up 2,000).

Finally, note that in this case, we do not need to standardize our parameters, as they represent mean group differences (see Section 2.4.1).

3.2.3 *Hypothesis Matrices*

Before computing the hypothesis matrix for each informative hypothesis, it is important to consider the contrasts' coding and the resulting parametrization of the encompassing model. We used default treatments contrasts for mother and father attachment (Schad et al., 2020), considering secure attachment as the reference category. Therefore, model intercept represents children with secure attachment toward both parents ($M_S F_S$) and we have parameters indicating the main effects of mother and father attachment and other parameters for the interaction effect.

Given the informative hypotheses and the parametrization of the encompassing model, we can obtain the respective hypotheses' matrices. To do that, we first need the model matrix with all the conditions of interest. Note that we only have to consider conditions relevant to the constraints (i.e., those related to mother and father attachment), ignoring other nuisance conditions (i.e., gender and classroom ID).

² Considering 1 as intercept (note that $\exp(1)$ is approximately the sample mean value), values included within 1 standard deviation, $\exp(1 \pm 1 \times SD)$, range between 0 and 55. Although externalizing problems are bounded between 0 and 20, prior predicted values are still reasonable, as they cover all possible values without including excessively large values. More diffuse priors would result in values with a higher order of magnitude, and tighter priors would exclude plausible values (see Supplemental Material for further details <https://claudiozandonella.github.io/Attachment/>)

Subsequently, we can derive the required equality and inequality constraints. In particular, with hypotheses that do not expect interaction between mother and father attachment (i.e., monotropy, hierarchy, and independence hypotheses), all interaction terms are set equal to zero, and main effects are obtained considering the reference level of the other parent (i.e., $M_{Ax}F_S$ is the main effect of anxious attachment toward the mother). As an example, consider the following code:

```
# Define relevant conditions
attachment <- c("S", "Ax", "Av", "F")
new_data <- expand.grid(
  mother = factor(attachment, levels = attachment),
  father = factor(attachment, levels = attachment)
)

# Get model matrix
mm <- model.matrix(~ mother * father, data = new_data)[,-1] # remove intercept
rownames(mm) <- paste0("M_", new_data$mother, "_F_", new_data$father)

# Get constraints main effect
# M_Ax = M_Av ----> M_Ax_F_S - M_Av_F_S = 0
# F_F > F_Av ----> M_S_F_F - M_S_F_Av > 0
rbind(mm["M_Ax_F_S", ] - mm["M_Av_F_S", ],
      mm["M_S_F_F", ] - mm["M_S_F_Av", ])

# Get constraints interaction
# M_F_F_Ax = M_Ax_F_F ----> M_F_F_Ax - M_Ax_F_F = 0
# M_Ax_F_Av > M_S_F_Ax ----> M_Ax_F_Av - M_S_F_Ax > 0
rbind(mm["M_F_F_Ax", ] - mm["M_Ax_F_F", ],
      mm["M_Ax_F_Av", ] - mm["M_S_F_Ax", ])
```

In this way, we can easily specify all the constraints obtaining the respective hypothesis matrix (R) and the vector with the constraints' constant values (r) for each hypothesis. In all our hypotheses, however, no constraints include constant values. Thus, r is always a vector of zeros and it can be ignored. Moreover, we can also ignore the intercept because no constraints include M_SF_S alone (e.g., $M_SF_S > k$); but it is always compared with other groups (e.g., $M_SF_S < M_{Ax}F_S$). Thus, the inclusion of the intercept is redundant and we can ignore it.

Full hypothesis matrix specification for each hypothesis is available in the Supplemental Material <https://claudiozandonella.github.io/Attachment/>.

3.2.4 *Computing the Bayes Factor*

So far, we defined the hypotheses' matrices, specified the encompassing prior, and obtained the model's posterior distribution. To compute the Bayes factor, we now need the adjusted prior and the posterior of the transformed parameters vector β (i.e., the parameters that identify the constraints) for each hypothesis.

- **Adjusted Prior β .** As reported in Section 2.2.3, the adjusted prior is required to properly evaluate the constraints. Applying Equation 10, we obtain the prior for the transformed parameter and then set the mean vector to zero.
- **Posterior β .** The same transformation used for the prior can be applied, this time considering the estimated posterior mean $\hat{\theta}$ and the estimated posterior covariance $\hat{\Sigma}_{\theta}$, to obtain the posterior distribution of the transformed parameters vector β .

As discussed in Section 2.4.3, Equation 10 requires hypothesis matrix R to be *full-row-rank* (i.e., all constraints must be linearly independent). However, this is not the case for the hierarchy hypothesis. Thus, we followed the solution presented in Section 2.4.3. Detailed information is available in the Supplemental Material <https://claudiozandonella.github.io/Attachment/>.

Now, we have all the elements required to compute the Bayes factor for each hypothesis as described in Section 2.3. Moreover, assuming that each hypothesis is equally likely a priori, we can calculate the posterior probability of each hypothesis as (Gu et al., 2014; Hoijtink, 2012):

$$\text{Posterior Probability } H_i = \frac{BF_{iu}}{\sum_i BF_{iu}}. \quad (29)$$

3.3 Results and Prior Sensitivity Analysis

Bayes factor and posterior probability of each hypothesis are reported in Table 3. Results clearly indicate that among the considered hypotheses, the Monotropy Hypothesis

is the most supported by the data.

Table 3

Bayes factor encompassing model and hypothesis posterior probabilities ($n_{subj} = 847$).

Hypothesis	Bayes Factor	Posterior Probability	
Null	2.4e+11	0.01	○
Monotropy	2.5e+13	0.98	●
Hierarchy	2.9e+11	0.01	○
Independence	4.0e+09	0.00	○
Integration	3.3e+09	0.00	○

Remember, however, that prior specification affects the Bayes factor results. It is therefore recommended, to evaluate the results we would obtain using different prior settings as well. In particular, we consider as possible priors for the parameters of interest:

- $\mathcal{N}(0, .5)$: Unreasonable tight prior
- $\mathcal{N}(0, 1)$: Tighter prior
- $\mathcal{N}(0, 3)$: Original prior
- $\mathcal{N}(0, 5)$: More diffuse prior
- $\mathcal{N}(0, 10)$: Unreasonably diffuse prior

The results of the prior sensitivity analysis are reported in Table 4.

Overall results consistently indicate the Monotropy Hypothesis as the most supported by the data. However, we can observe two distinct patterns. As the prior becomes more diffuse, the order of magnitude of the Bayes factor comparing each hypothesis with the encompassing model increases. Moreover, the probability of the Null Hypothesis increases with more diffuse priors, whereas the probabilities of the Hierarchy, Independence and Integration Hypothesis increases with tighter priors.

Table 4

Bayes factor encompassing model and hypothesis posterior probabilities (PP) under different prior settings ($n_{subj} = 847$).

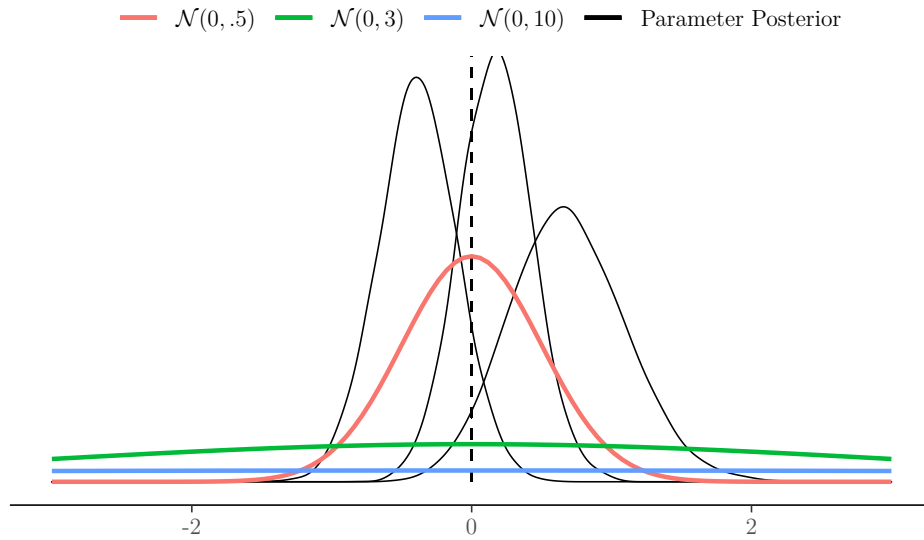
Hypothesis	$\mathcal{N}(0, .5)$		$\mathcal{N}(0, 1)$		$\mathcal{N}(0, 3)$		$\mathcal{N}(0, 5)$		$\mathcal{N}(0, 10)$	
	BF	PP	BF	PP	BF	PP	BF	PP	BF	PP
Null	8.9e+01	0.00	9.0e+04	0.00	2.4e+11	0.01	5.2e+14	0.03	1.4e+19	0.09
Monotropy	1.1e+05	0.67	6.6e+07	0.90	2.5e+13	0.98	1.7e+16	0.97	1.3e+20	0.90
Hierarchy	4.5e+04	0.28	6.5e+06	0.09	2.9e+11	0.01	6.8e+13	0.00	1.3e+17	0.00
Independence	4.0e+03	0.02	2.8e+05	0.00	4.0e+09	0.00	5.8e+11	0.00	5.6e+14	0.00
Integration	4.1e+03	0.03	3.5e+05	0.00	3.3e+09	0.00	3.3e+11	0.00	1.6e+14	0.00

To interpret these patterns, remember that the bayes factor is insensitive to the prior distribution specification for the parameters with order constraints as long as the distribution is symmetric and centered on the constraint focal point. On the contrary, in the case of equality constraints, the bayes factor is highly affected by the prior definition. As the prior gets more diffuse, the density value at zero decreases as well and, even for posterior distributions centered far from zero, the densities ratio at zero will favor the posterior. Thus, as the prior becomes more diffuse, the Bayes factor will increasingly favor the hypotheses with equality constraints. For tighter priors, the Bayes factor will strongly penalize hypotheses with equality constraints if these are not correct (see Figure 9).

All the defined hypotheses include equality constraints. Thus, for more diffuse priors we observe that the order of magnitude of the Bayes factor comparing each hypothesis with the encompassing model increases. Moreover, the hypothesis with the highest number of equality constraints (e.g., Null Hypothesis) will be favored over hypotheses with a smaller number of equality constraints (e.g., Hierarchy, Independence and Integration Hypothesis).

3.3.1 Limitations

When interpreting the results, it is important to take into account the limitations of the Bayes factor. First of all, the selected hypothesis is relative to the data and the set of hypotheses considered. That means that we should not interpret the selected hypothesis as the only “correct” one. Perhaps we did not consider some important aspects, and new

**Figure 9**

Evaluating densities at 0 for different prior settings and a selection of parameters posteriors.

hypotheses may reveal themselves to be actually much better than the previous ones. For example, regarding the attachment results, taking into account the interaction between children's gender and attachment may be fundamental (is it possible that attachment toward the father plays a different role between girls and boys?). This limitation, however, becomes an advantage as it forces the researchers to focus on the definition of the hypotheses: to obtain good answers, first we need to ask the right questions.

Moreover, considering a unique single winning hypothesis may be limiting, as different hypotheses could help explain different aspects of the process under investigation. For example, mother and father attachment may play different roles at different ages. Thus, the selected hypothesis may vary according to children's age or other conditions. This should warn against interpreting the results in terms of winning or losing, focusing uniquely on the selected hypothesis and discarding other hypotheses from further investigations.

Another important limitation we observed is the sensitivity of the results to the prior specification. Gelman et al. (2013) do not recommend using the Bayes factor, given

its sensitivity to model definition arbitrary choices. Evaluating the results under different conditions and transparently discussing the reasons behind arbitrary choices would help to interpret results more consciously (Schad, Nicenboim, et al., 2021).

Finally, another limitation related to the use of the Bayes factor with the encompassing prior approach is that we do not obtain the actual estimates of the parameters posterior. The only information we get is the selected hypothesis, but we have no other information regarding the actual parameters' values. Therefore, for example, we cannot assess the actual effect sizes. To overcome this limitation, we have to obtain via Bayesian inference the actual posterior of the model parameters given the prior, formalized according to parameters' constraints, and the observed data (for an introduction to Bayesian estimation Kruschke & Liddell, 2018). However, posterior estimation under inequality constraints requires the use of appropriate advanced computational algorithms (see Ghosal & Ghosh, 2022; Kato & Hoijtink, 2006).

4 Conclusion

In this paper, we presented a detailed description and an applied example of the Bayes factor with the encompassing prior approach. As discussed in the introduction, this approach has several advantages over traditional NHST: it is possible to evaluate complex hypotheses with equality and inequality constraints, multiple hypotheses can be tested simultaneously, and Bayes factor and posterior probabilities allow quantifying the relative evidence from the data in favor of the hypotheses.

This approach is of particular interest to the researchers, because informative hypotheses allow them to formalize their expectations with great flexibility. In this paper, we considered only the presence/absence of an effect (i.e., $\theta = 0$) or the expected effects order (i.e., $\theta_i > \theta_j$). However, in case of clear indications about the actual effects, researchers can specify the expected range of values for the effects of interest (i.e., $a < \theta_i < b$). This allows even a greater level of precision in the formalization of informative hypotheses.

As emphasized by Gu et al. (2018, p.229), “This class of informative hypotheses covers a much broader range of scientific expectations than the class of standard null hypotheses. In addition, by testing competing informative hypotheses directly against each other a researcher obtains a direct answer as to which scientific theory is most supported by the data”.

The available literature, however, is often technical and difficult to follow for non-experts in this approach. This paper filled this gap by providing a detailed description of all steps and elements involved in the computation of the Bayes factor with the encompassing prior approach. This will enhance researchers’ awareness regarding all pros and cons of this method, helping them avoid applying statistical techniques as black boxes.

Author Contributions

Conceptualization: C. Zandonella Callegher, T. Marci, P. De Carli, and G. Altoè; Methodology: C. Zandonella Callegher, T. Marci, P. De Carli, and G. Altoè; Software: C. Zandonella Callegher and G. Altoè; Formal Analysis: C. Zandonella Callegher and G. Altoè; Investigation: T. Marci and P. De Carli; Data Curation: T. Marci and C. Zandonella Callegher; Writing – Original Draft: C. Zandonella Callegher; Writing – Review & Editing: T. Marci, P. De Carli, and G. Altoè; Visualization: C. Zandonella Callegher; Supervision: G. Altoè; Project Administration: G. Altoè;

Conflicts of Interest

The authors declare that there were no conflicts of interest with respect to the authorship or the publication of this article.

Supplemental Material

All materials and data are available at <https://doi.org/10.17605/OSF.IO/JKMDF>, GitHub repository <https://github.com/ClaudioZandonella/Attachment>. Online supplemental material available via GitHub Pages at <https://claudiozandonella.github.io/Attachment/>.

Note that data used for the illustrative application are part of a larger project on the social-emotional development of school-age children, which was approved by the Ethics Committee of the School of Psychology at the University of Padova (protocol #[1838-2016]).

Prior versions

Preprint version of the submitted manuscript is available at <https://psyarxiv.com/6kc5u>.

References

- Ainsworth, M. D., & Bell, S. M. (1970). Attachment, exploration, and separation: Illustrated by the behavior of one-year-olds in a strange situation. *Child Development*, 41(1), 49–67. <https://doi.org/10.2307/1127388>
- Akaike, H. (1973, November). Information theory and an extension of the maximum likelihood principle. In B. Petrov & F. Csaki (Eds.), *Proceedings of the Second International Symposium on Information Theory* (pp. 267–281). Budapest: Akademiai Kiado.
- Altinisik, Y., Van Lissa, C. J., Hoijsink, H., Oldehinkel, A. J., & Kuiper, R. M. (2021). Evaluation of inequality constrained hypotheses using a generalization of the AIC. *Psychological Methods*, 26(5), 599–621. <https://doi.org/10.1037/met0000406>
- Anraku, K. (1999). An information criterion for parameters under a simple order restriction. *Biometrika*, 86(1), 141–152. <https://doi.org/10.1093/biomet/86.1.141>
- Bartlett, M. (1957). A comment on D. V. Lindley’s statistical paradox. *Biometrika*, 44, 533–534.
- Bayarri, M. J., & García-Donato, G. (2007). Extending conventional priors for testing general hypotheses in linear models. *Biometrika*, 94(1), 135–152. <https://doi.org/10.1093/biomet/asm014>
- Berger, J. O., & Pericchi, L. R. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, 91(433), 109–122. <https://doi.org/10.1080/01621459.1996.10476668>
- Bowlby, J. (1969). *Attachment and loss*. Basic Books.
- Brenning, K., Petegem, S. V., Vanhalst, J., & Soenens, B. (2014). The psychometric qualities of a short version of the Experiences in Close Relationships Scale – Revised Child version. *Personality and Individual Differences*, 68, 118–123. <https://doi.org/10.1016/j.paid.2014.04.005>

- Bretherton, I. (2010). Fathers in attachment theory and research: A review. *Early Child Development and Care*, 180(1-2), 9–23. <https://doi.org/10.1080/03004430903414661>
- Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, 10(1), 395. <https://doi.org/10.32614/RJ-2018-017>
- Cassidy, J., & Shaver, P. R. (2016). *Handbook of attachment: Theory, research, and clinical applications* (Third Edition). Guilford Press.
- de Santis, F., & Spezzaferri, F. (1999). Methods for default and robust Bayesian model comparison: The fractional Bayes factor approach. *International Statistical Review*, 67(3), 267–286. <https://doi.org/10.2307/1403706>
- Dickey, J. M. (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Mathematical Statistics*, 42(1), 204–223. <https://doi.org/10.1214/aoms/1177693507>
- Du, H., Edwards, M. C., & Zhang, Z. (2019). Bayes factor in one-sample tests of means with a sensitivity analysis: A discussion of separate prior distributions. *Behavior Research Methods*, 51(5), 1998–2021. <https://doi.org/10.3758/s13428-019-01262-w>
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC press. <https://doi.org/10.1201/b16018>
- Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.-C., & Modrák, M. (2020, November 3). *Bayesian workflow*. <http://arxiv.org/abs/2011.01808>
- Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6), 721–741. <https://doi.org/10.1109/TPAMI.1984.4767596>

- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2021). *mvtnorm: Multivariate normal and t distributions*.
<https://CRAN.R-project.org/package=mvtnorm>
- Ghosal, R., & Ghosh, S. K. (2022). Bayesian inference for generalized linear model with linear inequality constraints. *Computational Statistics and Data Analysis*, 17.
<https://doi.org/10.1016/j.csda.2021.107335>
- Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): Data from British parents, teachers and children. *Journal of Abnormal Child Psychology*, 38(8), 1179–1191.
<https://doi.org/10.1007/s10802-010-9434-x>
- Gu, X., Hoijtink, H., Mulder, J., & van Lissa, C. J. (2021). *Bain: Bayes factors for informative hypotheses*. manual. <https://CRAN.R-project.org/package=bain>
- Gu, X., Mulder, J., Deković, M., & Hoijtink, H. (2014). Bayesian evaluation of inequality constrained hypotheses. *Psychological Methods*, 19(4), 511–527.
<https://doi.org/10.1037/met0000017>
- Gu, X., Mulder, J., & Hoijtink, H. (2018). Approximated adjusted fractional Bayes factors: A general method for testing informative hypotheses. *British Journal of Mathematical and Statistical Psychology*, 71(2), 229–261.
<https://doi.org/10.1111/bmsp.12110>
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97–109. <https://doi.org/10.1093/biomet/57.1.97>
- Heck, D. W., Boehm, U., Böing-Messing, F., Bürkner, P. .-. C., Derks, K., Dienes, Z., Fu, Q., Gu, X., Karimova, D., Kiers, H., Klugkist, I., Kuiper, R. M., Lee, M. D., Leenders, R., Leplaa, H. J., Linde, M., Ly, A., Meijerink-Bosman, M., Moerbeek, M., ... Hoijtink, H. (2020, October 20). *A review of applications of the*

- Bayes factor in psychological research* (preprint). PsyArXiv.
<https://doi.org/10.31234/osf.io/cu43g>
- Heck, D. W., & Davis-Stober, C. P. (2019). Multinomial models with linear inequality constraints: Overview and improvements of computational methods for Bayesian inference. *Journal of Mathematical Psychology*, *91*, 70–87.
<https://doi.org/10.1016/j.jmp.2019.03.004>
- Hoijtink, H. (2012). *Informative hypotheses: Theory and practice for behavioral and social scientists*. Chapman and Hall. <https://doi.org/10.1201/b11158>
- Hoijtink, H., Mulder, J., van Lissa, C., & Gu, X. (2019). A tutorial on testing hypotheses using the Bayes factor. *Psychological Methods*, *24*(5), 539–556.
<https://doi.org/10.1037/met0000201>
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.
- Kato, B. S., & Hoijtink, H. (2006). A Bayesian approach to inequality constrained linear mixed models: Estimation and model selection. *Statistical Modelling*, *6*(3), 231–249.
<https://doi.org/10.1191/1471082X06st119oa>
- Kerns, K. A., & Richardson, R. A. (Eds.). (2005). *Attachment in middle childhood*. The Guilford Press.
- Klugkist, I., & Hoijtink, H. (2007). The Bayes factor for inequality and about equality constrained models. *Computational Statistics & Data Analysis*, *51*(12), 6367–6379.
<https://doi.org/10.1016/j.csda.2007.01.024>
- Klugkist, I., Kato, B., & Hoijtink, H. (2005). Bayesian model selection using encompassing priors. *Statistica Neerlandica*, *59*(1), 57–69.
<https://doi.org/10.1111/j.1467-9574.2005.00279.x>
- Klugkist, I., Laudy, O., & Hoijtink, H. (2005). Inequality Constrained Analysis of Variance: A Bayesian Approach. *Psychological Methods*, *10*(4), 477–493.
<https://doi.org/10.1037/1082-989X.10.4.477>

- Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, 25(1), 178–206.
<https://doi.org/10.3758/s13423-016-1221-4>
- Kuiper, R. M., Hoijsink, H., & Silvapulle, M. J. (2011). An Akaike-type information criterion for model selection under inequality constraints. *Biometrika*, 98(2), 495–501. <https://doi.org/10.1093/biomet/asr002>
- Kuiper, R. (2022). AIC-type Theory-Based Model Selection for Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(1), 151–158.
<https://doi.org/10.1080/10705511.2020.1836967>
- Kuiper, R., Hoijsink, H., & Silvapulle, M. (2012). Generalization of the order-restricted information criterion for multivariate normal linear models. *Journal of Statistical Planning and Inference*, 142(8), 2454–2463.
<https://doi.org/10.1016/j.jspi.2012.03.007>
- Lindley, D. V. (1957). A statistical paradox. *Biometrika*, 44(1-2), 187–192.
<https://doi.org/10.1093/biomet/44.1-2.187>
- Lindley, D. V. (1972). *Bayesian statistics: A review*. Society for Industrial and Applied Mathematics.
- Marci, T., Moscardino, U., & Altoè, G. (2019). The brief Experiences in Close Relationships Scale - Revised Child version (ECR-RC): Factor structure and invariance across middle childhood and early adolescence. *International Journal of Behavioral Development*, 43(5), 409–423.
<https://doi.org/10.1177/0165025418785975>
- McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (2nd ed.). Taylor and Francis.
- Mikulincer, M., & Shaver, P. R. (2007). *Attachment in adulthood: Structure, dynamics, and change*. Guilford Press.

- Mulder, J., & Olsson-Collentine, A. (2019). Simple Bayesian testing of scientific expectations in linear regression models. *Behavior Research Methods*, 51(3), 1117–1130. <https://doi.org/10.3758/s13428-018-01196-9>
- Mulder, J., & Raftery, A. E. (2022). BIC Extensions for Order-constrained Model Selection. *Sociological Methods & Research*, 51(2), 471–498. <https://doi.org/10.1177/0049124119882459>
- Mulder, J. (2014). Prior adjusted default Bayes factors for testing (in)equality constrained hypotheses. *Computational Statistics & Data Analysis*, 71, 448–463. <https://doi.org/10.1016/j.csda.2013.07.017>
- Mulder, J. (2016). Bayes factors for testing order-constrained hypotheses on correlations. *Journal of Mathematical Psychology*, 72, 104–115. <https://doi.org/10.1016/j.jmp.2014.09.004>
- Mulder, J., & Gelissen, J. P. T. M. (2019). *Bayes factor testing of equality and order constraints on measures of association in social research*. <http://arxiv.org/abs/1807.05819>
- Mulder, J., Hoijsink, H., & de Leeuw, C. (2012). BIEMS: A Fortran 90 program for calculating Bayes factors for inequality and equality constrained models. *Journal of Statistical Software*, 46(2). <https://doi.org/10.18637/jss.v046.i02>
- Mulder, J., & Wagenmakers, E.-J. (2016). Editors' introduction to the special issue "Bayes factors for testing hypotheses in psychological research: Practical relevance and new developments". *Journal of Mathematical Psychology*, 72, 1–5. <https://doi.org/10.1016/j.jmp.2016.01.002>
- Mulder, J., Williams, D. R., Gu, X., Tomarken, A., Böing-Messing, F., Olsson-Collentine, A., Meijerink, M., Menke, J., van Aert, R., Fox, J.-P., Hoijsink, H., Rosseel, Y., Wagenmakers, E.-J., & van Lissa, C. (2021). BFpack : Flexible Bayes Factor Testing of Scientific Theories in R. *Journal of Statistical Software*, 100(18). <https://doi.org/10.18637/jss.v100.i18>

- O'Hagan, A. (1995). Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society, Series B (Methodological)*, 57(1), 99–138.
- R Core Team. (2021). *R: A language and environment for statistical computing*. manual. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>
- Schad, D. J., Betancourt, M., & Vasishth, S. (2021). Toward a principled Bayesian workflow in cognitive science. *Psychological Methods*, 26(1), 103–126.
<https://doi.org/10.1037/met0000275>
- Schad, D. J., Nicenboim, B., Bürkner, P.-C., Betancourt, M., & Vasishth, S. (2021). *Workflow techniques for the robust use of Bayes factors*.
<http://arxiv.org/abs/2103.08744>
- Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, 110, 104038. <https://doi.org/10.1016/j.jml.2019.104038>
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- Stan Development Team. (2020). RStan: The R interface to Stan. <http://mc-stan.org/>
- van de Schoot, R., Mulder, J., Hoijtink, H., Van Aken, M. A. G., Semon Dubas, J., Orobio de Castro, B., Meeus, W., & Romeijn, J.-W. (2011). An introduction to Bayesian model selection for evaluating informative hypotheses. *European Journal of Developmental Psychology*, 8(6), 713–729.
<https://doi.org/10.1080/17405629.2011.621799>
- Varadhan, R. (2020). *condMVNorm: Conditional multivariate normal distribution*.
<https://CRAN.R-project.org/package=condMVNorm>
- Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11(1), 192–196.
<https://doi.org/10.3758/BF03206482>

- Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, 60(3), 158–189.
<https://doi.org/10.1016/j.cogpsych.2009.12.001>
- Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond “ $p < 0.05$ ”. *American Statistician*, 73, 1–19.
<https://doi.org/10.1080/00031305.2019.1583913>
- Wetzels, R., Grasman, R. P., & Wagenmakers, E.-J. (2010). An encompassing prior generalization of the Savage–Dickey density ratio. *Computational Statistics & Data Analysis*, 54(9), 2094–2102. <https://doi.org/10.1016/j.csda.2010.03.016>
- Zellner, A., & Siow, A. (1980). Posterior odds ratios for selected regression hypotheses. M. H. , D. V. Lindley, & A. F. M. Smith (Eds.), In *Bayesian statistics: Proceedings of the First International Meeting held in Valencia (Spain)* (pp. 585–603). University of Valencia Press.