Special Issue:
Bayesian Probability and Statistics in Management Research

Journal of Management
Vol. 41 No. 2, February 2015 544–573
DOI: 10.1177/0149206314525206

Relative Effects at Work: Bayes Factors for Order Hypotheses

Johan Braeken
Wageningen University and Research Centre
Joris Mulder
Tilburg University
Stephen Wood
University of Leicester

Assessing the relative importance of predictors has been of historical importance in a variety of disciplines including management, medicine, economics, and psychology. When approaching hypotheses on the relative ordering of the magnitude of predicted effects (e.g., the effects of discrimination from managers and coworkers are larger than that from clients), one quickly runs into problems within a traditional frequentist framework. Null hypothesis significance testing does not allow researchers to directly map research hypotheses on to results and suffers from a multiple testing problem that leads to low statistical power. Furthermore, all traditional structural equation modeling fit indices lose much of their suitability for model comparison, because order hypotheses are not countable in terms of degrees of freedom. To adequately tackle order hypotheses, we advocate a Bayesian method that provides a single internally consistent solution for estimation and inference. The key element in the proposed model comparison approach is the use of the Bayes factor and the incorporation of order constraints by means of a smart

Acknowledgments: The authors would like to acknowledge the editors, two anonymous reviewers, Sussane Beijer, Kevin Daniels, George Michaelides, Sandra Nolte, and Marc van Veldhoven for their feedback on earlier versions of the manuscript. The empirical data used are from the Service Delivery and Organisation (SDO) National Inpatient Staff Morale Study, which was funded by the National Institute for Health Research Service Delivery and Organisation (NIHR SDO) programme (Project Number 08/1604/142) and of which Stephen Wood was a co-investigator. We are grateful to Sonia Johnson, Chris Stride, and the participants in the study for their contribution to the Staff Morale Study. The views and opinions expressed in the article are those of the authors and do not necessarily reflect those of the NIHR SDO programme or the Department of Health. Joris Mulder's contribution to this article was financially supported by a Rubicon Grant provided by the Netherlands Organisation for Scientific Research (NWO).

Corresponding author: Johan Braeken, Faculty of Educational Sciences, Centre for Educational Measurement (CEMO), University of Oslo, P.O.Box 1161, Blindern, N-0318 Oslo, Norway.

E-mail: johan.braeken@cemo.uio.no

formulation of prior distributions. An easy-to-use software package BIEMS (Bayesian inequality and equality constrained model selection) is introduced and two empirical examples in the organizational behavior area are provided to showcase the method, both offering new findings that have implications for theory: the first on the differential impact of discrimination in the workplace from insiders and outsiders to the organization on employees' well-being, and the second on Karasek's stressor–strain theory about how the relative order of magnitude of the effects of job control and demands depends on the specific well-being outcome dimension.

Keywords: *Bayes factor; order hypotheses; model comparison; workplace discrimination; well-being; Karasek's job control-demands model*

When formulating research hypotheses about a phenomenon, we can hypothesize the mere presence or absence of a certain effect. Yet such null hypotheses have been increasingly questioned as they bear little risk of falsification and do not contribute much new information toward the further development of theories in a field. Edwards and Berry (2010) make this point explicit and call for a program of *strong inference* in organizational and management research. This program is to be based upon the formulation of more elaborate and precise theories (Edwards & Berry, 2010) and the development and testing of the corresponding informative hypotheses (see, e.g., Hoijtink, 2011). One approach towards this strong inference objective is to build hypotheses that stipulate a relative rank order for predicted effects in terms of their expected magnitude.

An example of such an *order hypothesis* can be found in the area of work-life balance. Both work-to-family and family-to-work types of interference can have antecedents and consequences in both the family and the work domain. If we focus on the consequences, competing theories suggest, on the one hand, that the effects of each will be most felt in the domain receiving the conflict so work-to-family interference will have the most effect in the family domain, for example family satisfaction will be reduced (Frone, Russell, & Cooper, 1992); on the other hand, the alternative posits that the effects will be felt most in the domain from which the conflict arises. A meta-analysis by Amstad, Meier, Fasel, Elfering, and Semmer (2011) testing these two theories, the cross-domain and matching theories, offered the strongest support for matching theory. Moreover, the results suggest that the issue is not simply which is right but rather which exhibits the stronger effects, since there is some evidence of cross-domain effects as well. Testing between the two theories is not then simply a matter of finding which effects are "significant" but involves a comparison of the differing theoretical orderings of predicted effects of both types of interference across the two outcome domains.

Yet in practice, people tend to talk about the relative importance of predictors in a rather vague descriptive way. However, comparisons of the sort "effect A is statistically significant, but effect B is not" can be highly misleading, because the difference between significant and nonsignificant is not guaranteed to be itself statistically significant (Gelman & Stern, 2006). For correct inference on the relative size of effects, we should test the effects relative to each other (i.e., A versus B), instead of separately testing the effects against zero, which is usually implied (i.e., A versus 0 and B versus 0). Nothing is large or small on its own, but such a statement always requires relative comparison; we should be looking at *relative effects*. Some work, with dominance analysis (Budescu, 1993) and relative weights (Johnson, 2000) as the most prominent examples, has been done on further quantification of the relative importance of an effect aimed at the development of a standardized effect size metric. However, it is impossible to find a unique solution to allocating explained variance across predictors such that it clearly quantifies their importance, and there are different ways of understanding "effects" as such. Whereas there is a whole literature that focuses on quantifying the relative importance of a predictor (for a review, see Johnson & LeBreton, 2004), the issue of statistical inference and hypothesis testing in the area of order hypotheses and comparisons of relative effect orderings has been neglected.

The effect is that the traditional frequentist framework that currently dominates in management research and elsewhere has as yet no convenient statistical tools to approach such order hypotheses. Traditional null hypothesis significance testing (NHST) does not allow researchers to directly map complex research hypotheses onto results and traditional model comparison has no way of incorporating order constraints into its fit indices (Mulder et al., 2009). To adequately tackle order hypotheses, we advocate a Bayesian method that provides a single internally consistent solution for estimation and inference. The key element in the proposed model comparison approach is the use of the Bayes factor (Jeffreys, 1961; Kass & Raftery, 1995) and the incorporation of order constraints by means of a smart formulation of prior distributions.

Before we proceed to the core of the article, the introduction of this Bayesian approach to order hypotheses, we will first discuss the kinds of order hypotheses that are of interest in the field of management and then highlight the problems that one encounters when dealing with order hypotheses within the traditional frequentist framework. Then, to ensure that the article is self-contained, we will start with an outline of the key concepts, terminology, and ideas underlying Bayesian data analysis[1] and continue with a more detailed explanation on the use of Bayes factors for statistical inference on order hypotheses.

The method will be illustrated through two empirical examples in the organizational behavior area. The first example involves the impact of discrimination in the workplace from different sources on employees' well-being, in which the sources (e.g., managers, coworkers, patients, visitors) are ordered according to the relative power of their position within the organization and are expected to differentially affect well-being accordingly (i.e., univariate order hypothesis on the relative magnitude of predicted effects). The second example involves different interpretations of the Karasek theory of psychological strain (Karasek, 1979) in which the relative impact of its key variables, job control and demands, may depend on the specific well-being dimension, leading to relative orderings that differ across outcomes (i.e., multivariate order hypotheses on the relative magnitude of predicted effects within and across outcomes). The examples illustrate that the more traditional fit indices are not informative for choosing among competing order hypotheses, whereas the Bayes factor allows for an informative assessment and quantification of support for these hypotheses and has the benefit of offering a more intuitive and probability-oriented interpretation compared to traditional model fit indices. For both illustrative investigations, we use data from a large survey of mental health workers in England (Johnson et al., 2012). To promote and support the practical use of this innovative Bayesian method, we have developed a freely available software package BIEMS (Bayesian inequality and equality constrained model selection) (with user-friendly interface) that computes these Bayes factors between models with order (and other) constraints on the parameters of the multivariate linear model.

Order Hypotheses in Management: Relevance and Examples

An order hypothesis stipulates a relative rank order for predicted effects in terms of their expected magnitude. There are many instances where theories imply such an ordering or where alternative theories might be identified on the basis of different orderings. Comparative rank orderings can be derived in several ways and plenty of opportunities can be found in the field of management.

First, a relative rank ordering can be hypothesized based upon how close predictors are to the outcome in the causal flow. For example, in the study of psychological contracts (Rousseau, 1995), a relational breach of the psychological contract by a proximal source (e.g., a line manager) can be expected to have a worse effect on an employee than would a transactional breach from a distal source (e.g., a senior manager). Second, a logical ordering may already be present in the phenomenon at hand. The quality of leader-member exchange, for instance, may systematically depend on the leader's level within the organization relative to the subordinate, or on the complexity of the tasks involved. Third, competing theories may give priority to different theoretical components. For example, in the high performance work systems area, the ordering of performance effects of the different practices associated with the high performance model may vary depending on one's underlying theory (Posthuma, Campion, Masimova, & Campion, 2013; Wood, 2013).

We use the term *univariate order hypothesis* to refer to hypotheses that make comparisons between predictors of the same outcome and the term *multivariate order hypothesis* to refer to hypotheses that make comparisons between predictors across different outcomes. We have already seen such a multivariate example in the area of work-life balance with the competing matching-domains and cross-domains theories. A second multivariate example is the long-standing debate about the relative performance of formality versus informality on a range of outcomes in, for instance, the areas of employee relations, innovation, and training (Heyes, 2001). Formal structures are expected in employment relations to have more impact on performance than informal voice (Bryson, Willman, Gomez, & Kretschmer, 2013) and training policy is largely predicated on the value of formal methods, though the emphasis in the recent debate on workplace learning is on the potential, perhaps equal or greater, value of informal methods (Eraut, 2004; Griffin, 2011), or in contrast, in the case of innovation, formalization and bureaucracy may be an impediment (Adler & Borys, 1996; Thompson, 1965).

Finally, widening the scope even more to multigroup order hypotheses, cross-cultural comparison is another relevant area where order might be of central interest when one has theories on the nature of structural or measurement invariance (Vandenberg & Lance, 2000) between countries or organizations. For example, the effect of trade unionism on wages at the workplace level may be greater in countries with a decentralized system than in countries with a coordinated economy or mixed economy, or the rank ordering of optimal human resource strategies might differ between individualistic and collectivistic cultures, or an anticipated ordering of differential item functioning between groups or across cultures in employment-selection instruments.

Compared to overly general omnibus questions or null hypotheses, focused research questions such as order hypotheses have greater conceptual clarity and yield greater statistical power. If the hypothesis is supported, we are more likely to discover it and to believe it to be real when asking focused questions rather than general ones (Rosenthal, Rosnow, & Rubin, 2000). Furthermore, theories that survive increasingly stringent order hypotheses will contribute to a stronger theoretical foundation (Edwards & Berry, 2010). However, we also need appropriate statistical tools that are able to answer such focused research questions and test the corresponding order hypotheses. Before we outline the Bayesian approach to order hypotheses, we illustrate the problems one encounters within the classical frequentist framework when trying to deal with order hypotheses, such that it is clear why an alternative approach is needed.

The Frequentist Approach to Order Hypotheses

As a working example to help our theoretical exposition, consider the case of workplace aggression, a phenomenon which may have profound negative effects on workers' well-being (Bowling & Beehr, 2006). In the workplace aggression literature it is suggested that, compared with aggression from outsiders, aggression originating from inside the organization (e.g., supervisors or coworkers) may have more negative consequences (see, e.g., Hershcovis & Barling, 2010). This has triggered research that compares the effect of aggression from different sources inside and outside the organization. In a health-care setting, for instance, we can think of sources such as managers (M), coworkers (C), patients (P), and visitors (V). In the context of a regression model, this gives rise to the following regression equation:

$$\textit{Well-Being} = \beta_{\mathrm{M}}\mathrm{Agression}(\mathrm{M}) + \beta_{\mathrm{C}}\mathrm{Agression}(\mathrm{C}) + \beta_{\mathrm{P}}\mathrm{Agression}(\mathrm{P}) + \beta_{\mathrm{V}}\mathrm{Agression}(\mathrm{V}),$$

where the β's are the regression effects of the various aggression sources on well-being. The power differential in the manager-subordinate relationship suggests that there may be a stronger negative impact of aggression from managers (Aquino, Tripp, & Bies, 2001) compared to aggression from others. Expanding this power differential principle, we might also assume that workplace aggression by visitors has the least negative effect because visitors are outsiders entering the organization's premises. We expect the middle ground to be covered by coworkers and patients. Coworkers are insiders having both organizational rights and obligations towards patients. The patients, though outsiders, have some power over staff. Thus, the research hypothesis of interest involves a comparison between four sources—managers (M), coworkers (C), patients (P), and visitors (V)—and the expected ordering of effects is based upon power differentials between the different roles in the settings of a mental health organization. We operate under the expectation that workplace aggression, regardless of the source, has a negative effect (i.e., all betas smaller than zero) on the victim's well-being. Symbolically this order hypothesis can be written as:

$$\mathrm{H}: \beta_{\mathrm{M}} < (\beta_{\mathrm{C}}, \beta_{\mathrm{P}}) < \beta_{\mathrm{V}} < 0. \quad (1)$$

NHST. In a case like the order hypothesis in Equation 1, traditional NHST breaks down, as it requires a tremendous number of significance tests that are not necessarily independent or unambiguous. First, four tests are needed for each individual effect, then multiple pairwise comparison tests of the effects need to be conducted to determine if they are significantly different from each other and in the expected direction. Subsequently, it will be very difficult, if not impossible, to combine the resulting set of p values into one single answer to the question of whether the order hypothesis is supported by the data. Also, multiple testing requires type I error corrections (e.g., Bonferroni alpha-correction), which will result in very conservative tests with very little power (Cohen, 1992). Another issue is that there is no straightforward way to understand what the null hypothesis is in this context. For example, do you compare all effects to each other, or only those that are supposed to be closest in magnitude? How do you deal with contradictory results in pairwise comparisons (e.g., β_M is significantly different from β_C, but not from β_P, yet β_C is not significantly different from β_P)? Basically, with order hypotheses or any other moderately complex set of hypotheses, traditional NHST is going to be ineffective. It fails to directly assess the research question of interest.

Traditional model comparison. Instead of NHST, a model comparison approach can be followed as in the popular structural equation modeling (SEM) literature (see, e.g., Bollen, 1989), but even this will not produce definitive assessments of order hypotheses. In such an approach, we start with an a priori set of theoretical models, and then assess and compare these models as a function of the deviation of the sample data from the model and the parsimony of the model. The model that compared to competing models offers an equally good but less complex explanation of the data will be the working model of choice. Balancing fit and parsimony provides guarantees for generalization of the model results to other samples (see, e.g., Hastie, Tibshirani, & Friedman, 2001; Pitt & Myung, 2002). In SEM, model selection is based upon fit statistics (see, e.g., Hu & Bentler, 1999) such as the root mean squared error of approximation (RMSEA) or comparative fit index (CFI), that are formalized according to this "Ockham's razor" principle. Deviation between data and model is expressed in terms of a Chi-square measure (χ^2: summarizing the compatibility of the observed and model-implied covariance matrix) and parsimony is expressed in terms of degrees of freedom (df: the difference between the number of known sufficient sample statistics and the number of unknown model parameters):

$$\begin{aligned}\text{Fit} &= \text{f}\left(\text{Deviation, Parsimony}\right)\\ &= \text{f}\left(\chi^2,\ \text{df}\right).\end{aligned} \tag{2}$$

In our example case of workplace aggression, we would tailor the statistical model such that it corresponds as closely as possible to our a priori expectation as formulated in the order hypothesis of Equation 1. Most popular SEM software allow researchers to add the necessary constraints on the model by putting restrictions on the possible values the regression coefficients can take. Hence, an order hypothesis as in Equation 1 can be directly translated into an order-constrained statistical model. If these constraints are quite unreasonable, the constrained model will necessarily show large deviations from the data. If the constraints are consistent with the data, the constrained model will give an approximately equally good yet more parsimonious account of the data patterns than the unconstrained model. Unfortunately, traditional model fit indices are unsuitable for selecting between competing order-constrained models, a point which we will now explain theoretically, but which we will also demonstrate in practice in the first empirical application, the discrimination case (see Table 3, discussed later in this article).

To explain where the incompatibility between traditional fit indices and order hypotheses comes from, let us start with the case of a simple null hypothesis. If we expect aggression from managers and coworkers to have no effect on well-being, we could model this by constraining their effects to be equal to zero. The unconstrained alternative model is obviously less parsimonious than this constrained model, as the former has the flexibility to be consistent with any possible data pattern. The constrained model sets $\beta_M = \beta_C = 0$ and thus has two parameters less, and hence two more degrees of freedom than the unconstrained model. There is a one-to-one correspondence between theoretical precision, model parsimony, and degrees of freedom in the case of equality constraints. In contrast, by imposing the order constraint that the effect of managers is lower in value than the effect of coworkers (i.e., $\beta_M < \beta_C$), we do increase the precision of our theory and the informativeness of our hypothesis, but we do not reduce the number of regression coefficients (model parameters), and hence, the degrees of freedom of the constrained model remains exactly the same as those of the unconstrained alternative. Nevertheless, as with equality constraints, order constraints limit the model to a smaller range of potential data patterns, making it more parsimonious and less flexible. Yet simply because the constraints are on the possible parameter values these regression coefficients can take and do not imply a concrete reduction of model parameters, the logic of operationalizing theoretical precision and model parsimony in terms of countable degrees of freedom breaks down. This means that in case of order hypotheses, traditional fit indices reduce to simple deviation measures (see Equation 2) and do not result in a reward for formulating a more focused informative hypothesis. An overly general unconstrained model that fits would be assessed to be equally good as an a priori more risky (i.e., in terms of fit) but more specific and informative order-constrained model. This is highly undesirable if we are to achieve Edwards and Berry's aim of a strong inference and theoretical precision in management.

Thus, to summarize, both traditional NHST approaches as well as those based on model fit as in SEM are not suitable to test theories and hypotheses involving order constraints. Therefore, we need an alternative approach that is able to test order hypotheses. In the next two sections we will present a Bayesian approach that fits this purpose.

Bayesian Data Analysis

In this section, we introduce key terminology and concepts in a Bayesian approach with respect to (a) model specification and (b) model comparison. We build upon this introduction to explain how the Bayes factor, a Bayesian model comparison tool, can be used for proper statistical inference on order hypotheses.

Model Specification: Prior, Likelihood, and Posterior

The design of a research study is typically informed by theory and prior results from the literature. Bayesian data analysis builds upon this starting point by formalizing these prior expectations in terms of probability distributions for the parameters of interest in a statistical model. Hence, such a *prior distribution* reflects the information we have about the model parameters before observing the data. This can be based on external information sources such as expert judgment and subjective theory, or more empirically based on meta-analysis results of the existing research literature. Another option is to deliberately remain very uncertain and formulate a so-called noninformative (or *vague*) prior distribution. For the effects of discrimination on well-being, for instance, a uniform prior between an extremely negative and extremely positive value (say between -100 and 100) could be specified. In this case every possible effect across this range is equally likely a priori. This is conceptually similar to taking no theoretical perspective before the analysis starts, and for instance disregarding that the effect will almost certainly be negative. We will denote the prior distribution of the effects β_t under model M_t as $\Pr(\beta_t|M_t)$, where the subscript t simply indicates a number assigned to refer to a specific statistical model or hypothesis (the sign "|" should be read as "conditional on" or "under").

Having formulated a prior distribution for the parameters of the model, the next step in Bayesian data analysis is to *update* the prior expectations with the information provided by the collected sample data. This information is formalized in terms of the *likelihood* which links the data to the statistical model. Note that the likelihood has the same role as in classical statistics. We denote a likelihood by $\Pr(\text{data}|\beta_t, M_t)$, which reads as the probability of the observed sample data under a fixed value β_t for the parameter of model M_t.

The update, combining information from the prior and from the data, results in the so-called *posterior distribution*, denoted by "$\Pr(\beta_t|data, M_t)$," which quantifies the plausibility of the effects β_t under model M_t after observing the data. This posterior distribution can be used for traditional forms of estimation and hypothesis testing about parameters (Kruschke, Aguinis, & Joo, 2012): derivation of point estimates (e.g., posterior mode or posterior mean), credibility intervals (i.e., the Bayesian equivalent of confidence intervals), and hypothesis tests for single model parameters or even for nonlinear combinations of model parameters.

Computationally, the underlying updating mechanism that defines Bayesian analysis is based on an application of Bayes' theorem and often formulated as:

$$\underbrace{\Pr(\beta_t|data, M_t)}_{Posterior} \propto \underbrace{\Pr(\beta_t|M_t)}_{Prior} \times \underbrace{\Pr(data|\beta_t, M_t)}_{Likelihood}, \tag{3}$$

where Equation 3 reads as "*posterior is proportional to prior times likelihood.*"

In contrast to a Bayesian approach, a classical frequentist data analysis approach would seek to find the point estimate for the parameter, denoted by $\hat{\beta}_t$, for which the likelihood is maximized. Prior distributions are not used in a frequentist analysis, and hence one is always in an uninformed prior setting and only informed by the data. The downside is that one only has a single value for the model parameters and not the broader probability information available in a Bayesian approach. Yet for some the frequentist approach still sounds like a more objective and less controversial approach than a Bayesian data analysis with priors. However, the actual impact of a prior distribution decreases proportionally with increasing sample size (i.e., the data will dominate the prior) and a Bayesian approach can always mimic a frequentist analysis by using noninformative priors.

The reverse cannot be said; often for complex hypothesis tests or more complex statistical models, no straightforward frequentist solution exists, when a Bayesian approach can offer a viable solution. Where a frequentist approach fails, the smart use of priors in a Bayesian approach will allow us to construct a proper tool for model comparison among competing models that reflect different order hypotheses. For this reason, one should not think of prior specification as an Achilles' heel of Bayesian statistics. Instead, priors and Bayesian data analysis in general provide researchers with a very flexible and broadly applicable informative approach to statistical inference.

Model Comparison: Marginal Likelihoods and Bayes Factors

Above we introduced the idea of Bayesian data analysis and estimation, where prior expectations are updated by the likelihood, a mechanism that was summarized in Equation 3. The exact version of the Bayes Theorem, given in Equation 4 below,

$$\underbrace{\Pr\left(\beta_t|\text{data}, M_t\right)}_{\textit{Posterior}} = \frac{\overbrace{\Pr\left(\beta_t|M_t\right)}^{\textit{Prior}} \times \overbrace{\Pr\left(\text{data}|\beta_t, M_t\right)}^{\textit{Likelihood}}}{\underbrace{\Pr\left(\text{data}|M_t\right)}_{\textit{Marginal likelihood}}} \tag{4}$$

includes yet another term that is commonly referred to as the marginal likelihood, denoted by $\Pr\left(\text{data}|M_t\right)$. This term can be omitted in estimation procedures because it is a normalizing constant that does not affect general Bayesian estimation techniques and is only theoretically necessary to make the posterior a "proper" probability distribution (i.e., with total probability summing to 1). However, for model comparison purposes, this marginal likelihood is a very informative measure.

Formally the marginal likelihood of a model M_t, $\Pr\left(\text{data}|M_t\right)$, is the expected probability of the data across the whole range of values of the prior distribution of the model parameters. In essence, the marginal likelihood computes an average level of fit to the data for a model M_t and summarizes the amount of evidence in the data for model M_t. More complex models that correspond to less specific theories can fit many different data patterns by changing their parameter values, but will have lower average levels of fit than simpler models that can fit a more limited set of data patterns and correspond to more specific theories (see, e.g., Lodewyckx et al., 2011). Raftery (1993) argues that the marginal likelihood is the optimal criterion for model comparison, as it incorporates and balances both model fit as well as model parsimony.

The Bayes factor is a Bayesian model comparison criterion that builds upon this balancing property of the marginal likelihood and quantifies the relative evidence between two scientific theories or hypotheses (Jeffreys, 1961; Kass & Raftery, 1995). The Bayes factor between two models, say M_1 and M_2, is defined as the ratio of the marginal likelihoods of the two models:

$$\text{BF}\left(M_1, M_2\right) = \frac{\Pr\left(\text{data}|M_1\right)}{\Pr\left(\text{data}|M_2\right)}. \tag{5}$$

In contrast to traditional fit indices, a Bayes factor is interpretable without recourse to qualifiers or commonly accepted thresholds (such as a significance level of .05). For instance, when the Bayes factor is larger than 1, say, BF(M_1, M_2) = 10, there is simply 10 times more evidence in the data for model M_1 than for M_2. When the Bayes factor is smaller than 1, say BF(M_1, M_2) = 0.10, model M_1 receives 10 times less evidence from the data than M_2. Because marginal likelihoods balance model fit and parsimony, the Bayes factor will favor the simpler explanation (i.e., model or hypothesis) that does adequately fit the data. Bayes factors have some convenient properties that add to an easy and consistent interpretation of supporting evidence in the data for the models that are being compared. Reversing the roles of the two models simply leads to reversing the Bayes factor: BF(M_1, M_2) = 1/ BF(M_2, M_1); or in other

Table 1
Guidelines for Interpreting Bayes Factors Following Kass and Raftery (1995)

BF(M_1,M_2)	Evidence in Favor of M_1
< 1	Negative (i.e., evidence for M_2)
1–3	Anecdotal
3–20	Positive
20–150	Strong
>150	Very strong

words, if there is four times more evidence in the data for model M_1 than for model M_2, this logically implies that there is four times less evidence in the data for model M_2 than for model M_1. Bayes factors between different models can also be indirectly linked, as long as they have a common referent using the transitivity property. For instance, if BF(M_1, M_2) = 4 and BF(M_3, M_2) = 8, then BF(M_3, M_1) = 8/4 = 2; or in other words, if model M_1 and M_3 have, respectively, four and eight times more evidence than model M_2, then this implies that model M_3 has two times more evidence than model M_1.

As with effect sizes (Cohen, 1992), tentative guidelines have been provided on how to interpret the outcome of the Bayes factor in a more absolute sense. Table 1 shows the interpretation as proposed by Kass and Raftery (1995). These guidelines can be used as a convenient default starting point for interpreting Bayes factors. Hence, researchers can make statements such as there is "positive" or "very strong" evidence in the data for a certain model against another model based on the observed Bayes factor.

Bayes Factors for Order Hypotheses

In this section, we outline the specific approach to implementing Bayes factors such that they can be used to test order hypotheses. In the general case, computing Bayes factors can be complex because we need to mathematically integrate (cf. "average") the likelihood function across possible values of the prior distribution for the model parameters. However, in some cases, convenient shortcut computations are possible. We will first present the logic behind our chosen approach, which is based on having an encompassing unconstrained model and then tackle somewhat more technical details on the specification of prior distribution that is required for the approach to work.

Throughout the section we use a working example that consists of three order-constrained regression models with two predictor variables "aggression from Managers" (M) and "aggression from Coworkers" (C) that affect the outcome variable depression:

$$M_1 : \beta_M > 0, \beta_C > 0$$

$$M_2 : \beta_M > \beta_C$$

$$M_3 : \beta_M > \beta_C > 0$$

$$M_u : \beta_M \text{ and } \beta_C \; (\textit{unconstrained}).$$

Model M_1 focuses on the signs of the effects and does not contain any relative ordering. The model states that the two regression parameters are positive, which implies that aggression by both managers and coworkers results in an increase in depression. Model M_2 does stipulate an ordering on the magnitude of predicted effects: the effect of aggression from managers (β_M) is larger than the effect of aggression from coworkers (β_C). Model M_3 is a combination of M_1 and M_2: aggression by managers and coworkers both result in an increase in depression, but aggression by managers has a stronger effect. Finally, the unconstrained model M_u that makes no assumptions at all about the effects of aggression on depression is included.

An Encompassing Model Approach to Bayes Factors

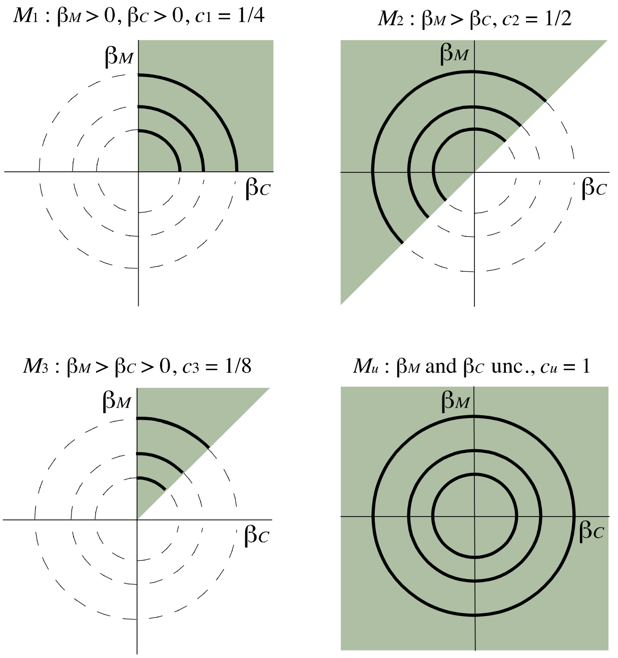
Making use of the proportionality of order-constrained models to an unconstrained model, the computation of Bayes factors for order hypotheses can avoid complicated integration problems. The only requirement is that an *encompassing* unconstrained model exists, which is the more general version of the order-constrained model(s) so that all models are nested within the encompassing model. The shortcut computation then comes down to determining the Bayes factor between a constrained model M_{t} and the unconstrained model M_{u} as the posterior probability that the order constraints hold under the unconstrained model (i.e., f_t a measure of fit for M_t) weighted by the prior probability that the order constraints hold under the unconstrained model (i.e., c_t a measure of complexity for M_t) (for the technical derivation of this shortcut, see the Appendix):

$$\mathrm{BF}\left(M_t, M_u\right) = \frac{f_t}{c_t}. \tag{6}$$

We will first discuss the two elements in the ratio of Equation 6, prior complexity and posterior fit, and then show how the combination of both yields a Bayes factor that can quantify the evidence in the data for an order hypothesis and in turn facilitates straightforward comparisons between different order hypotheses. For the latter purpose, we will show how the concept of posterior model probability enables clear and easy interpretation of the model comparison results.

Prior complexity. Figure 1 illustrates the proportionality idea for the prior distributions and the complexity term c_t. Each model in the working example is represented in a bidimensional plot where the *x*-axis represents the possible values the effect of aggression from coworkers β_C can be expected to take a priori under the unconstrained model, and where the *y*-axis has a similar function but for the effect of aggression from managers β_M. This bidimensional plot is the graphical representation of the prior distribution of β_M and β_C under the unconstrained model. The gray areas represent the combination of values of β_M and β_C that satisfy the constraints of each individual model. The complexity of the unconstrained model is for instance equal to 1 as all values for β_M and β_C are possible, which means that the whole area is covered (Figure 1, lower right panel). For Model 1 (Figure 1, upper left panel), only positive values of β_M and β_C are possible, constraining the area to the upper-right quadrant that takes up one fourth of the space compared to the unconstrained model. The complexity of Model 1 is therefore equal to one fourth, less complex than the unconstrained

Figure 1
Prior Complexity c_t of Order-Constrained Models

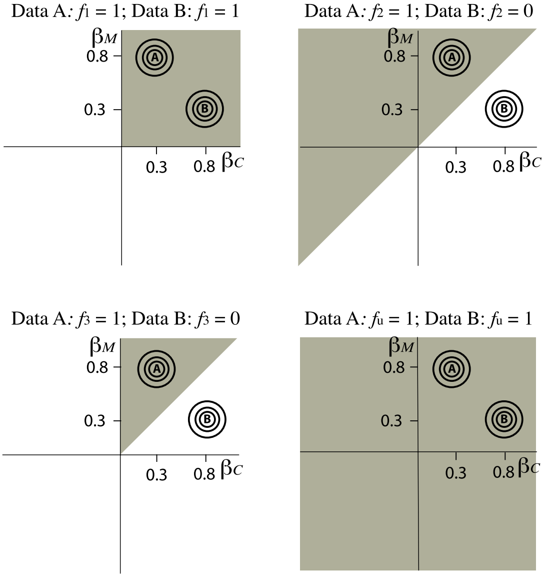


Note: The gray areas show which combinations of values of β_C and β_M satisfy the constraints of each model. The circular solid contours illustrate which values are most likely a priori (within the small circle) and least likely (outside the large circle). The truncated priors under the order-constrained are displayed as truncated solid circles of the unconstrained prior (the conjugate expected constrained posterior prior, CECPP) under M_u, with the dashed parts resembling zero probability areas.

model and only compatible with a limited set of possible results. In a similar fashion, for Model 2 (Figure 1, upper right panel) only half of the area is consistent with the constraint that $\beta_M > \beta_C$, and its complexity is consequently equal to a half. Because M_3 is a combination of M_1 and M_2, the constrained area of M_3 is equal to the area where the constrained areas of M_1 and M_2 overlap, which comes down to a complexity of one eighth (Figure 1, lower left panel). The model that corresponds to the most informative hypothesis/precise theory results in the most specific constrained area and the lowest complexity measure.

Posterior fit. To illustrate the proportionality idea for the posterior distribution and the fit term f_t, we will consider two hypothetical data sets under the same set of models: Data A with maximum likelihood estimates $\hat{\beta}_M = 0.8$ and $\hat{\beta}_C = 0.3$ and Data B with maximum likelihood estimates $\hat{\beta}_M = 0.3$ and $\hat{\beta}_C = 0.8$ (both under the unconstrained model). Figure 2 locates the posterior distribution of β_M and β_C under the unconstrained model for Data A and Data B as smaller inner circles in the same plots as we used for the prior distributions. The fit measure in use is a direct reflection of the extent to which the posterior distributions fall into the area of the unconstrained prior distribution that is consistent with the order constraints of the model. For Data A, for instance, we can conclude that the posterior falls completely within

Figure 2
Posterior Fit f_t for Data A and Data B of Order-Constrained Models



Note: The gray areas show which combinations of values of β_C and β_M satisfy the a priori constraints of each model. The circular solid contours illustrate which values are most likely a posteriori for Data A and Data B.

the area with possible values that are consistent with the constraints within each model. This implies that the fit of all order-constrained models is maximal and that the model fit terms are all equal to 1. For Data B, we conclude that model M_1 shows maximal fit equal to 1, but that models M_2 and M_3 show complete misfit with a fit measure equal to zero because the posterior distribution does not overlap with the a priori possible area.

Bayes factor. Balancing the complexity and the fit measure as in Equation 6 then leads to the Bayes factor (i.e., BF(M_t, M_u) = f_t / c_t). As under Ockham's razor, when models equally fit the data (which is the case for Data A, see Table 2), the least complex model (i.e., M_3) is preferred—in this case, with the Bayes factor indicating that, compared to the unconstrained model M_u, there is eight times more evidence in the data for model M_3, which is two times more than for M_2 and four times more than for model M_1. For Data B we can see that there is four, zero, and zero times more evidence for M_1, M_2, and M_3, respectively, than for the unconstrained model M_u. Hence, based on these Bayes factors, models M_2 and M_3 receive no support at all from the data. This also makes sense because the order constraints are not supported by the maximum likelihood estimates based on Data B.

Bayes Factors and Posterior Model Probabilities

In this section, we will connect the Bayes factor to what is called posterior model probability. The latter will be useful for interpretation of the results of comparisons between

Table 2
Bayes Factors (BFs) Between Each Order-Constrained Model and the Unconstrained Model and Corresponding Posterior Model Probabilities (PMPs) for Data A and Data B

	Data A				Data B			
	c_t	f_t	BF(M_t,M_u)	PMP	c_t	f_t	BF(M_t,M_u)	PMP
M_1	1/4	1	4	0.27	1/4	1	4	0.80
M_2	1/2	1	2	0.13	1/2	0	0	0.00
M_3	1/8	1	8	0.53	1/8	0	0	0.00
M_u	1	1	1	0.07	1	1	1	0.20

multiple models, but also for illustrating that Bayes factors are not affected by common misconceptions that haunt frequentist hypothesis testing and p values.

Similar to the mechanism that we described to update the prior with the likelihood to obtain the posterior, the Bayes factor can be used to update the prior model odds to obtain the posterior model odds according to:

$$\underbrace{\frac{\Pr(M_1 \mid \text{data})}{\Pr(M_2 \mid \text{data})}}_{\text{Posterior model odds}} = \underbrace{\text{BF}(M_1, M_2)}_{\text{Bayes factor}} \times \underbrace{\frac{\Pr(M_1)}{\Pr(M_2)}}_{\text{Prior model odds}} . \quad (7)$$

In Equation 7, $\Pr(M_t)$ denotes the prior model probability of M_t, which is the probability that model M_t is true before observing the data, and $\Pr(M_t \mid \text{data})$ denotes the posterior model probability of M_t, the probability that model M_t is true after observing the data. Note that the prior model probability $\Pr(M_t)$ is different from the prior distribution $\Pr(\beta_t|M_t)$: the first specifies how likely model M_t is before observing the data, while the latter specifies how large the effects β_t (i.e., model parameters) are expected to be under model M_t a priori. Because in practice prior model probabilities are often chosen to be equal (i.e., $\Pr(M_1) = \Pr(M_2)$), the prior model odds in Equation 7 reduce to 1 and the posterior odds are then simply equal to the Bayes factor.

Posterior odds for a set of models can be transformed into a set of posterior model probability for these models. Consider the four models in our working example, assuming equal prior model probabilities of $\Pr(M_1) = \Pr(M_2) = \Pr(M_3) = \Pr(M_u) = 1/4$, we can use the Bayes factors of all models against the same benchmark model to compute a set of posterior model probabilities. The posterior model probability of a model M_t can be obtained as a normalized weight:

$$\Pr(M_t|\text{data}) = \frac{\text{BF}(M_t, M_u)}{\text{BF}(M_1, M_u) + \text{BF}(M_2, M_u) + \text{BF}(M_3, M_u) + \text{BF}(M_u, M_u)}, \quad (8)$$

(i.e., a proportion computed as the Bayes factor of one model in the set against the reference model divided by the sum of Bayes factors of all models in the set against the same reference). The resulting posterior model probabilities for Data A and Data B can be found in Table 2. For Data A, model M_3 has the largest posterior model probability of 0.53, followed by gradual decrease over M_1 and M_2, towards M_u, and for Data B, model M_1 has clearly the largest posterior model probability of 0.80, followed by M_u, whereas models M_2 and M_3 result in posterior model probabilities of zero. Posterior model probabilities always will sum up to 1 within a set, offering a convenient frame of reference for the relative magnitude of a posterior probability. In our experience researchers find that this reference frame makes posterior model probabilities easier to interpret compared to Bayes factors, where they might need to rely on guidelines with which they are not familiar. The risk is, of course, that people might be inclined to request fixed cut-off values for posterior model probabilities in a similar fashion to the widely used 0.05 p value threshold in the case of NHST. But the Bayes factors and posterior model probabilities are best used as relative measures of evidence between two or more models, and they avoid some of the pitfalls of the traditional p values.

With the concept of posterior model probability we can understand better some misconceptions that surround classical p values. In fact classical p values are often misinterpreted as being the (posterior model) probability that a null model is true, but a p value is only equal to the probability of observing more extreme data than the observed data under the condition that the null hypothesis is true. As shown by Sellke, Bayarri, and Berger (2001), the misinterpretation of the p value as a posterior model probability results in serious overestimation of the evidence against the null model. Under very reasonable assumptions (such as equal prior model probabilities), a p value of 0.05 results in a posterior model probability that the null model is true of 0.29, showing how p values underestimate evidence in favor of null hypotheses. Furthermore, whereas we might be inclined to assume that a significant model M_1 with $p < 0.01$ is 10 times more supported than a significant model M_2 with $p < 0.001$, we would be making a serious error of judgment and confuse p values with error probabilities (Hubbard & Armstrong, 2006). In contrast, the ratio of two posterior model probabilities 0.01 versus 0.001 (computed among the same set of models) is simply equal to the corresponding Bayes factor $\text{BF}(M_1, M_2) = 0.01/0.001 = 10$. These discrepancies illustrate that researchers should be very careful when interpreting p values as measures of evidence for model selection and hypothesis testing, and that posterior model probabilities and Bayes factors are to be preferred over classical p values because of their more intuitive interpretation.

Details on Prior Specification

This subsection will deal with the practical implementation of our approach. The contents will therefore be particularly useful for readers who are more familiar with the technical aspects of statistical modeling.

Prior specification. When specifying the priors for the effects β_M and β_C, it is important to ensure that the order-constrained models are *nested* in the unconstrained model M_u. This allows one to specify a so-called encompassing prior under the unconstrained model, that is subsequently truncated to obtain priors under the order-constrained models (e.g., Hoijtink, 2011; Klugkist, Kato, & Hoijtink, 2005). Hence, in our working example we would only need to decide on how to specify the encompassing prior for the effects β_M and β_C under the unconstrained model. This encompassing prior is chosen to have a bivariate normal distribution, which is directly compatible (i.e., a *conjugate* form) with the normal linear model (see, e.g., Press, 2005). We choose a bivariate normal prior that is centered and symmetrical around zero. In this case there is no prior preference towards either a positive or a negative effect and no prior preference whether the effect of aggression from managers β_M is larger or smaller than the effect of aggression from coworkers β_C. This prior can therefore be considered to be "objective" or impartial (i.e., showing no prior preference for a specific hypothesis) with respect to hypothesis tests of order constraints on relative effects in linear models (Mulder, Hoijtink, & Klugkist, 2010).

Subsequently, the variances of the encompassing prior must be specified. To avoid arbitrary or ad hoc specification of the prior variance, random subsets of minimal size are selected from the data, so-called minimal training samples, to determine how large relative effects may deviate from the prior means of zero (Berger & Pericchi, 2004). These deviations are then averaged to obtain the prior variances. Although it may seem counterintuitive to use subsets of the data for specifying the prior, the method can best be seen as a formal way to obtain a relatively vague prior that does not contradict the data. The resulting encompassing prior is also referred to as the *conjugate expected constrained posterior prior* (CECPP; Mulder et al., 2009, 2010; Mulder, Hoijtink, & de Leeuw, 2012). The priors under the order-constrained models are then simple truncations of the CECPP in a fashion similar to that presented in Figure 1 with zero prior probability in areas that are inconsistent with the order constraints.

Software Package BIEMS

BIEMS (Bayesian inequality and equality constrained model selection; Mulder et al., 2012) is a freely downloadable software package (www.jorismulder.com) with a user-friendly interface that implements the proposed Bayesian approach to hypothesis testing and model comparison. The two unique aspects of BIEMS are the automatically generated CECPP prior distribution, which ensures no prior preference towards a certain ordering of the effects, and the efficient computation of Bayes factors through the balancing of model fit f_t and model complexity c_t.

To use BIEMS, first, the data must be uploaded and the number of outcome variables and predictors must be specified. Second, the models must be specified with order constraints and equality constraints. Third, the automatic CECPP prior is constructed. In the fourth optional step, one could view and if necessary alter the prior (e.g., when a priori knowledge is available). In the fifth and final step, Bayes factors of each constrained model against the unconstrained model and other model diagnostics are reported. A user manual can be found in Mulder et al. (2012). An accessible tutorial paper about the use of BIEMS is provided by Kluytmans, van de Schoot, Mulder, and Hoijtink (2012).

Despite the effectiveness of Bayesian data analysis, Bayesian methods have not been implemented in many commercial statistical packages, and for now, Bayesian testing of order-constrained models has not yet been implemented elsewhere. Recent developments in Mplus show some promise for the future as the development team decided to implement Bayesian estimation methods because of their superiority over frequentist methods in the case of small samples and its flexibility in fitting complex models using Markov chain Monte Carlo (MCMC) techniques (Muthén & Asparouhov, 2012). The latest version of Mplus supports a very broad Bayesian estimation approach, but unfortunately a formulation of the necessary CECCP priors is not yet possible. We expect that future versions will allow the user to have more control over the priors enabling the method proposed here. For the time being, van de Schoot, Hoijtink, Hallquist, and Boelen (2012) have suggested a very rudimentary pragmatic approach by manually computing the model fit term f_t as the proportion of posterior simulations under the unconstrained model that are in line with the order hypothesis and approximating the model complexity term c_t by means of a proportion based upon the number of possible rank orderings (permutations) among predictors. This might be a little bit too involved for regular use, and it is difficult to apply it to hypotheses with many order constraints.

The following section illustrates our approach in practice and discusses two empirical examples from organizational research where BIEMS was effectively used to determine which of a set of scientific theories receives most evidence from the data.

Empirical Examples

In our two empirical examples we use data from a large survey conducted amongst mental health workers (Johnson et al., 2012). The outcomes in both examples focus on the well-being of the workers. Well-being was operationalized by complementary three-item scales that measure anxiety and depression (Warr, 1990), and a five-item scale measuring job dissatisfaction (a combination scale based on the UK's 2004 Workplace Employment Relations Survey [National Institute of Economic and Social Research, 2004] and the National Staff Survey [National Health Service, 2006]). The first example draws on a study by Wood, Braeken, and Niven (2013) on perceived workplace discrimination and involves assessing the relative impact of discrimination in the workplace from different sources on employees' well-being (i.e., managers, coworkers, patients, visitors). The second example involves Karasek's (1979) theory of psychological strain and is on different interpretations about how the relative order of magnitude of the effects of job control and demands depends on the specific well-being outcome dimension. To illustrate the range of the Bayes factor approach, the first example will test univariate hypotheses involving an ordering of predictors for one outcome, and the second example multivariate hypotheses involving mixed orderings of the same predictors across different outcomes. To ensure meaningful comparison of regression effects in each example, variables were measured on the same scale within each set of predictors, and the outcome variables were also on a similar scale for ease of interpretation.

In both examples, the order-constrained models are first tested against the unconstrained model using the Bayes factor. This is a first check to determine whether the model is suitable for describing the data. A Bayes factor larger than one for a constrained model against the unconstrained model implies that there is more evidence for the constrained model. The guidelines in Table 1 offer useful assistance in the assessment of the strength of support for a specific model as indicated by the Bayes factor. Subsequently for further model comparison purposes, posterior model probabilities are computed for all models under the assumption that all models are equally likely a priori. To illustrate the practical implementation of order-constrained models in the software package BIEMS, we indicate the set of separate constraints that need to be specified to build up each model. We do not necessarily posit all possible alternative models, but concentrate on fairly simple models that connect to theory in the two areas.

Application 1: Workplace Discrimination From Inside and Outside the Organization

In the first example we assess competing hypotheses on the effect of perceived discrimination from different sources inside and outside the organization. Considerable evidence exists on detrimental psychological and health effects of experiencing discrimination at work (Goldman, Gutek, Stein, & Lewis, 2006). Perceived workplace discrimination is operationalized by means of single-item measures that capture individuals' own summary terms they use to reflect their feelings (Wood, Stride, & Johnson, 2012: 538). It captures people's perceptions of whether they have been discriminated against, as opposed to imposing a label, which victims might not agree with based on various behaviors that may or may not have been interpreted in an adverse manner. Because the study was aimed at a broad range of issues and not all issues could be covered in-depth, the use of single-item measures also reflected the practical constraints for the study with respect to survey length and cognitive load. Respondents simply indicated whether they experienced any form of discrimination in the past 12 months with reference to each of the following potential sources: managers, coworkers, patients, and visitors. Thus, we have instances of workplace discrimination from four different sources inside and outside the organization.

Model 1: Negative impact (M_1). The first model is rather unspecific and only makes the assumption that perceived discrimination decreases a worker's well-being regardless of the source. Given that our well-being outcomes are all negatively formulated—anxiety, depression, job dissatisfaction—we expect all regression coefficients of the discrimination sources to be positive (i.e., discrimination implies more anxiety). This constrained hypothesis is formulated as:

$$M_1 : (\beta_M, \beta_C, \beta_P, \beta_V) > 0 \Leftrightarrow \begin{cases} \beta_M > 0 \\ \beta_C > 0 \\ \beta_P > 0 \\ \beta_V > 0 \end{cases},$$

where we denote after the left curly bracket the separate restrictions that need to be specified in the BIEMS software to construct this model. Table 3 provides the order constraints of this model, and its competitor models below, as implemented in BIEMS.

Model 2: Power differential (M_2). The second model builds on the insider-outsider theory of the effects of workplace aggression introduced earlier (Hershcovis & Barling, 2010), according to which the effects of discrimination from insiders, because they are assumed to have more power over victims than do outsiders, will have more effect on victims' well-being. Elaborating on this power-differential hypothesis, we expect that discrimination from one's manager will have the largest impact on the employee. In line with this reasoning, we assume that workplace discrimination by visitors has the least negative effect, because visitors are outsiders entering the organization's premises. Within the settings of a mental health organization, the middle ground is expected to be covered by coworkers and patients. Coworkers are insiders who may have high expertise and referent power, the latter being

Table 3
Order-Constrained Models as Implemented in BIEMS for Application 1

Model	Order Constraints in BIEMS
M_1 (negative impact):	α (1,1) > 0, α (2,1) > 0, α (3,1) > 0, α (4,1) > 0
M_2 (power diff.):	α (1,1) > α (2,1), α (1,1) > α (3,1), α (2,1) > α (4,1), α (3,1) > α (4,1), α (4,1) > 0
M_3 (con. sov.):	α (1,1) > α (3,1), α (3,1) > α (2,1), α (3,1) > α (4,1), α (2,1) > 0, α (4,1) > 0
M_4 (social contact:	α (2,1) > α (1,1), α (2,1) > α (4,1), α (3,1) > α (1,1), α (3,1) > α (4,1), α (1,1) > 0, α (4,1) > 0

Note: The regression coefficient of predictor *s* on outcome variable *t* is represented as α(*s,t*) in BIEMS. The regression coefficient corresponding to the effect of discrimination by managers, coworkers, patients, and visitors were the first, second, third, and fourth predictor, respectively.

particularly important as it is likely "to affect the presence and quality of social relationships within the group" (Hershcovis & Barling, 2010: 28). The patients are outsiders, but do have some power over staff and are both subject to organizational policies and users of organizational policies. This order hypothesis is formulated as:

$$M_2 : \beta_M > (\beta_C, \beta_P) > \beta_V > 0 \Leftrightarrow \begin{cases} \beta_M > \beta_C \\ \beta_M > \beta_P \\ \beta_C > \beta_V \\ \beta_P > \beta_V \\ \beta_V > 0 \end{cases}.$$

We again operate under the expectation that workplace discrimination, regardless of the source, is detrimental to the victim's well-being (i.e., corresponding to Model 1). The part ($\beta_{C,}$ β_P) indicates that we do not specify an ordering between these two coefficients. After the left curly bracket we show the separate restrictions that need to be specified in the BIEMS software to construct this model (see Table 3). It is not necessary to add, say, $\beta_M > \beta_V$, because this constraint is a direct logical consequence of $\beta_M > \beta_C$ and $\beta_C > \beta_V$, which are already specified.

Model 3: Consumer sovereignty and patient's rights (M_3). The third model acknowledges the primary position of the manager, but posits the patients as the second most impactful source of discrimination, having a greater effect than that of colleagues or visitors. An argument might be made that notions such as customer sovereignty and citizens/patients rights place an onus on representatives of organizations to treat customers, clients, and patients "with courtesy and as if they are right" (Grandey, Kern, & Frone, 2007: 65) even if those people are acting inappropriately. Thus, the customer, client, or patient is ordained with some degree of legitimate power. Moreover, patients and other outsiders, like customers (Gettman & Gelfand, 2007), also have some power to reward or punish staff, and how staff relates to patients is increasingly given prominence in appraisal and reward systems. Furthermore, in situations where the perpetrator can develop a more enduring relationship with the organization, as is the case with a patient who has a long hospital stay, referent power can be developed. As such, patients may have greater power compared with work colleagues, as well as visitors. Maintaining the assumption that all sources have an effect (see Model 1), this order hypothesis is then formulated as:

$$M_3 : \beta_M > \beta_P > (\beta_C, \beta_V) > 0 \Leftrightarrow \begin{cases} \beta_M > \beta_P \\ \beta_P > \beta_C \\ \beta_P > \beta_V \\ \beta_C > 0 \\ \beta_V > 0 \end{cases}.$$

Model 4: Social contact (M_4). The fourth hypothesis is based on the idea that the intensity of social contact with the different sources might be the key explanation behind their impact on a worker's well-being. Given that mental health workers in general have more frequent face-to-face contact with both colleagues and patients, discrimination from them can be hypothesized to have more impact on victim's well-being than that from managers or visitors. This order hypothesis is then formulated as:

$$M_4 : (\beta_C, \beta_P) > (\beta_M, \beta_V) > 0 \Leftrightarrow \begin{cases} \beta_C > \beta_M \\ \beta_C > \beta_V \\ \beta_P > \beta_M \\ \beta_P > \beta_M \\ \beta_M > 0 \\ \beta_V > 0 \end{cases}.$$

We again operate under the expectation that workplace discrimination, regardless of the source, is detrimental to the victim's well-being (i.e., corresponding to Model 1).

Model comparison. Bayes factors were computed for each of the four constrained models against the unconstrained model, using BIEMS. The results for the three outcomes—anxiety, depression, and job dissatisfaction—can be found in the upper panel of Table 4, together with the resulting posterior model probabilities and values for a popular traditional fit index, the Comparative Fit Index (CFI). In the lower panel of Table 4, the 95% credibility intervals of the regression coefficients are represented for the unconstrained model.

To illustrate problems that one might encounter when adopting a traditional model comparison approach to competing order-constrained models, we first look at the resulting CFI values for the different models. Based upon a CFI below the common acceptable fit threshold of 0.95, we can conclude that Model 4 does not show adequate fit to the data, but we also are unable to distinguish between the other models and this for all three outcome variables. They all appear to fit equally well (CFI's all equal to 0.99 or 1). This result provides us with little conclusive information to determine which of our hypotheses is most supported by the data. In contrast, Bayes factors do provide an informative way of selecting between competing order-constrained models by adequately balancing fit and parsimony.

Table 4 shows that for each outcome there is positive supporting evidence for Model 1 (as indicated by Bayes factor between 12.3 and 15.9), which is consistent with all discrimination sources having a negative impact on well-being. However, the Bayes factors of zero for

Table 4

Bayes Factors (BF) for the Constrained Models (M_t) Versus the Unconstrained Model (M_u), Posterior Model Probabilities (PMP), and 95% Credibility Intervals of the Regression Coefficients of the Effects of the Four Sources of Discrimination, Derived for Each Outcome Variable Separately

	Model Comparison Results								
	Anxiety			Depression			Job Dissatisfaction		
Model/Outcome	BF(M_t, M_u)	PMP	CFI	BF(M_t, M_u)	PMP	CFI	BF(M_t, M_u)	PMP	CFI
M_u	1.0	0.01	1	1.0	0.01	1	1.0	0.00	1
M_1 (negative impact)	14.7	0.08	1	15.9	0.16	1	12.3	0.06	1
M_2 (power diff.)	68.6	0.39	1	68.0	0.69	0.99	110.8	0.50	1
M_3 (con. sov.)	91.9	0.52	1	13.8	0.14	0.99	96.6	0.44	1
M_4 (social contact)	0.0	0.00	0.85	0.0	0.00	0.86	0.0	0.00	0.88

	95% Credibility Intervals for the β Effects Under M_u		
Predictors/Outcome	Anxiety	Depression	Job Dissatisfaction
Managers	[0.47, 0.80]	[0.54, 0.86]	[0.56, 0.89]
Coworkers	[0.01, 0.33]	[0.17, 0.49]	[0.06, 0.37]
Patients	[0.08, 0.41]	[0.01, 0.32]	[0.13, 0.45]
Visitors	[-0.06, 0.39]	[-0.00, 0.44]	[-0.14, 0.30]

[a]CFI is the comparative fit index, an instance of the traditional fit indices that are limited by their operationalization of complexity in terms of countable degrees of freedom.

Model 4 indicate that this model describes the data badly and its constraints do not hold, a result that is in agreement with the CFI-based conclusion and with the 95% credibility intervals of the regression coefficients in the lower panel of Table 4. The Bayesian analysis clearly shows that this social contact hypothesis (i.e., Model 4) underplays the most impactful source of discrimination in this sample, this being the managers. These results point in the direction of Models 2 and 3.

Based upon the Bayes factors we can conclude that Model 2 receives consistently strong support across the three outcomes, whereas Model 3 receives less support for depression (i.e., Bayes factor of 13.8 compared to Bayes factors of over 90 for anxiety and job satisfaction). When inspecting the posterior model probabilities, we observe that for depression Model 2 (the power differential hypothesis) is best supported (i.e., proportional weight of .69 in the set of models). But for both anxiety and job dissatisfaction Model 2 and Model 3 receive roughly equally strong support with posterior model probability of 0.39 versus 0.52 and 0.50 versus 0.44, respectively; this corresponds to about equal odds for both models as represented by Bayes factors between Model 2 and 3 of 0.75 and 1.13. What these two models share is that they put managers first and visitors last in the impact ordering of discrimination sources. There is not enough evidence to decide between the relative strength of the effect of discrimination from patients and that from coworkers on anxiety or job satisfaction.

Whereas a significance testing approach would force us to make a binary decision, the Bayes factor approach does not force us to choose between these two models; instead we can quantify how much weight or belief we attach to each model in the form of these posterior model probabilities. Bayesian methods give a clearer view of uncertainty in terms of probability and we should not shy away from such statements.

Based upon these sample data, we can conclude that there is no strong support for the social contact hypothesis and that we are to gauge between two hypotheses, the power differential and consumer sovereignty and patient's rights propositions. Further theoretical developments might want to concentrate on their mutual overlap, but also try to explore whether other measures of well-being yield a more clear differentiation between the two hypotheses (as is the case for depression).

Application 2: Interpretations of the Karasek Theory of Psychological Strain

Karasek's (1979) job strain model is highly influential in the occupational psychology and the wider management and work literature. It posits that psychological strain or stress is related to both the demands placed on an individual and their degree of job control or discretion. The core argument is typically based on energy depletion: demands place the individual in a motivated or energized state, while the degree of control modulates the transformation of potential energy into action. Accordingly, job demands are positively related to strain as they trigger it, and job control negatively as this enables the individual to more readily respond to the demands. As energy becomes depleted, strain intensifies.

Here we compare different hypotheses about how the relative order of magnitude of the effects of job control and demands depends on the specific well-being outcome dimension (e.g., the effect of demands overrules the effect of control for anxiety, but it's the other way around for their effects on job satisfaction). These are multivariate order hypotheses because effects are compared not only between predictors but also across outcomes, and the expected orderings between predictors are not consistent across the different outcomes. Order constraints will be formulated between six regression coefficients, $\beta_{X.Y}$, based on two different predictor X-variables—control (Con) and demands (Dem)—and three different outcome Y-variables—anxiety (Anx), depression (Dep), and job dissatisfaction (Jds). For example, $\beta_{Con.Anx}$ denotes the effect of control on anxiety.

Model 1: Directional effects model (M_1). The first interpretation of Karasek's theory that we investigate is limited to a priori expectations about the direction of the effects of control and demands. Because more control is expected to result in a positive effect on a worker's well-being and more demands is expected to result in a negative effect on a worker's well-being, the regression coefficients are expected to have the following signs:

$$\beta_{Con.Anx} < 0, \beta_{Dem.Anx} > 0,$$

$$\beta_{Con.Dep} < 0, \beta_{Dem.Dep} > 0,$$

$$\beta_{Con.Jds} < 0, \beta_{Dem.Jds} > 0.$$

Table 5
Order-Constrained Models as Implemented in BIEMS for Application 2

Model	Order Constraints in BIEMS
M_1: Directional effects model	
	$\alpha(1,1) < 0, \alpha(1,2) < 0, \alpha(1,3) < 0,$ $\alpha(2,1) > 0, \alpha(2,2) > 0, \alpha(2,3) > 0$
M_2: Warr's circumplex model	
	$-\alpha(1,1) < \alpha(2,1), -\alpha(1,2) > \alpha(2,2), -\alpha(1,3) > \alpha(2,3),$ $-\alpha(1,1) < -\alpha(1,2), -\alpha(1,1) < -\alpha(1,3),$ $\alpha(2,1) > \alpha(2,2), \alpha(2,1) > \alpha(2,3)$
M_3: Self-determination model	
	$-\alpha(1,1) > \alpha(2,1), -\alpha(1,2) > \alpha(2,2), -\alpha(1,3) > \alpha(2,3),$ $-\alpha(1,1) > -\alpha(1,3), -\alpha(1,2) > -\alpha(1,3)$

[a]Control is the first predictor, and demand is the second predictor. Anxiety, depression, and job dissatisfaction are the first, second, and third outcome variable, respectively.

Table 5 provides the order constraints of this model, and its two competitor models below, as implemented in BIEMS.

Model 2: Warr's circumplex model (M_2). The second interpretation contains the same constraints as Model 1, but adds a theoretical layer that builds on expectations that the ordering of the effects of demands and control will not be the same across the different dimensions of well-being. Following the circumplex theory of affect (Russell, 1980), we may differentiate feelings of anxiety from depression. Anxiety combines low pleasure with high arousal, while its opposite contentment is associated with higher levels of pleasure and low arousal. In contrast, depression involves feelings of low pleasure and low mental arousal, and enthusiasm entails feelings of high pleasure and high mental arousal. According to Warr (2007), job demands are more likely to generate anxiety than depression, whereas the frustration and loss of autonomy associated with jobs with low control is more likely to lead to depression. Job satisfaction is concerned only with the pleasure dimension and not arousal. Therefore, it can be anticipated to behave more like depression than anxiety, as the pleasure of work is depleted more by high demands than is the contentment. Thus, within each outcome variable, we expect the following relative magnitude ordering of effects:

$$-\beta_{Con.Anx} < \beta_{Dem.Anx},$$

$$-\beta_{Con.Dep} > \beta_{Dem.Dep},$$

$$-\beta_{Con.Jds} > \beta_{Dem.Jds}.$$

A minus is included in front of $\beta_{Con.Anx}$, $\beta_{Con.Dep}$, and $\beta_{Con.Jds}$ because these effects are expected to be negative. Across the three outcome variables we expect that the effects of control are strongest for job dissatisfaction compared to the other two outcomes, whereas the effects of demands are expected to be strongest for anxiety. This yields:

$$(-\beta_{Con.Anx}, -\beta_{Con.Dep}) < -\beta_{Con.Jds},$$

$$\beta_{Dem.Anx} > \left(\beta_{Dem.Dep}, \beta_{Dem.Jds}\right).$$

These last order constraints must be specified separately in BIEMS for each term within the brackets [e.g., $\beta_{Dem.Anx} > \left(\beta_{Dem.Dep}, \beta_{Dem.Jds}\right)$ should be implemented as $\beta_{Dem.Anx} > \beta_{Dem.Dep}$ and $\beta_{Dem.Anx} > \beta_{Dem.Jds}$] (see Table 6).

Model 3: Self-determination model (M_3). The third model contains the same constraints as Model 1 but adds different order constraints than Model 2. This interpretation is inspired by self-determination theory, according to which autonomy, along with competence and connectivity, are basic needs of workers (Deci & Ryan, 1995). Although demands may indirectly fulfill basic needs such as honing competence, control directly fulfills the basic need for autonomy, and as such we might hypothesize that control will have a greater effect on well-being than will demands:

$$-\beta_{Con.Anx} > \beta_{Dem.Anx},$$

$$-\beta_{Con.Dep} > \beta_{Dem.Dep},$$

$$-\beta_{Con.Jds} > \beta_{Dem.Jds}.$$

Although no specific ordering is expected regarding the relative effects of demands' across the three outcome variables, control is expected to have a stronger effect on anxiety and depression than on job dissatisfaction:

$$(-\beta_{Con.Anx}, -\beta_{Con.Dep}) > -\beta_{Con.Jds}.$$

Model comparison. Bayes factors were computed for each of the three constrained models against the unconstrained model, using BIEMS. The results are presented in the upper panel of Table 6, together with the resulting posterior model probabilities. In the lower panel of Table 6, the 95% credibility intervals of the regression coefficients are represented for the unconstrained model.

A result that stands out when inspecting the model comparison results is that Model 3, which was based upon self-determination theory, does not get any support, as evidenced in a Bayes factor of 0 (Bayes factors and posterior model probabilities of zero for M_3). When we look at the credibility intervals in the lower panel of Table 6, we can also observe that control is not always the dominant predictor as suggested by Model 3. In contrast, there is strong evidence for the directional effects of Model 1 against the unconstrained model as indicated by a Bayes factor of 63. This is in agreement with the 95% credibility intervals, which are

Table 6
Bayes Factors for the Constrained Models (M_t) Versus the Unconstrained Model (M_u), Posterior Model Probabilities (PMP), and 95% Credibility Intervals of the Regression Coefficients of the Effects of Job Control and Demands on Well-Being

	Model Comparison Results	
Model	BF(M_t, M_u)	PMP
M_1: Directional effects model	63	0.24
M_2: Circumplex-Warr model	194	0.76
M_3: Self-determination model	0	0.00
M_u: Unconstrained model	1	0.00

	95% Credibility Intervals for the β Effects Under M_u		
Predictors/Outcome	Anxiety	Depression	Job Dissatisfaction
Control	[-0.20, -0.12]	[-0.30, -0.22]	[-0.54, -0.47]
Demands	[0.27, 0.33]	[0.25, 0.32]	[0.16, 0.21]

located in the negative/positive area for the control/demand variable. This result is consistent with the expectation that more demands decrease well-being and more control increases well-being. When we look further at the more specific order hypotheses, the Bayes factor of 194 indicates that there is even stronger evidence for Warr's circumplex Model 2. Using the transitivity property, we can formally state that the evidence in favor of Model 2 is about three times (i.e., ratio of Bayes factors against the unconstrained model: 194/64 = 3.13) stronger than the evidence in favor of Model 1. Based on the posterior model probability of 0.76, it can be concluded that the circumplex model has the most support amongst our competitor models. The 95% credibility intervals of the regression coefficients show that most order constraints of M_2 are indeed supported by the data, except for $-\beta_{Con.Dep} > \beta_{Dem.Dep}$.

To follow up on this last observation in a more exploratory fashion, we evaluated the model without this particular constraint. This resulted in an increase in the Bayes factor against the unconstrained models to 618. This strong increase in support reinforces the results for anxiety and job satisfaction. A lack of control will have less effect on anxiety than will high demands, whereas this lack of control will have more impact on dissatisfaction than will high demands. Taken alongside the result for depression, the latter result suggests it is the pleasure dimension of well-being that is most differentially affected by control and demands. In contrast, the depression result suggests that the lack of control does not necessarily lead to any stronger decrease in arousal than does high demands, as Warr's theory specifies. Further research on this specific point and cross-validation of the results in other settings and samples might shed some more light on this and spark new theoretical developments.

Conclusion

This article has shown that a traditional frequentist approach is unable to adequately map complex research hypotheses onto informative results. In the case of order hypotheses, the key problem is the operationalization of model parsimony in terms of countable degrees of freedom. In the presence of order constraints, traditional fit indices (see, e.g., RMSEA, CFI, but also information criteria such as AIC or BIC) reduce to mere measures of deviation between model and data and lose their balancing system that penalizes complexity and rewards model parsimony.

In contrast, Bayes factors provide a flexible and broadly applicable Bayesian approach to model comparison. The use of prior information often allows one to bypass problems that exist in a frequentist approach. In the case of order hypotheses, we can connect the complexity of order-constrained models directly to the intuitive a priori relative size of the constrained area of possible effects (see, e.g., Figure 1). We demonstrated in the empirical examples that, compared to traditional fit indices, this allows Bayes factors (and posterior model probabilities) to provide a more insightful informative assessment of the relative support for complex hypotheses that put an ordering across predictors and/or outcomes (i.e., univariate and multivariate).

The use of Bayesian methods does not of course eliminate the need for cross-validation and further replication studies to take into account statistical and sample uncertainty, and consolidate supported theories. A big advantage of the Bayesian approach is that existing study results and evidence can be incorporated in the specification of priors (see, e.g., example 2 in Zyphur & Oswald, 2015) such that theories and research projects can progress incrementally as priors are updated by the results of new replication studies.

If we are to meet Edwards and Berry's (2010) call for greater theoretical rigor within the domain of management research, the theories will become more specific and research hypotheses will become more complex and involved. We can anticipate that this development will attract researchers to new and promising statistical tools for model comparison and hypothesis testing, and as we have shown, Bayes factors are likely to be at the forefront of this. Bayes factors are already gaining popularity as a model comparison tool in other social and life sciences areas such as cognitive psychology (Massaro, Cohen, Campbell, & Rodriguez, 2001), experimental psychology (Kammers, Mulder, de Vignemont, & Dijkerman, 2009), health-care studies (Spiegelhalter, Abrams, & Myles, 2004), and genetics (Stephens & Balding, 2009).

An increased exposure to Bayesian methods within management will, we assume, lead to applied researchers taking more interest in formulating more complex and informative research questions, such as ordered hypotheses, and using Bayesian methods to test these hypotheses. This increased interest may in turn encourage software development that makes Bayesian methods more practically accessible for use by applied researchers in the field. We hope that this article has gone some way towards stimulating these reciprocal effects.

Appendix: Origins of the Encompassing Model Shortcut

Here, we want to give some insight into the origins of the computational shortcut that can be used in the presence of an encompassing model (see also Klugkist et al., 2005) by showing how the original definition of the Bayes factor as the ratio of marginal likelihoods of M_t and M_u (as in Equation 5) can be expressed as the ratio of posterior fit and prior complexity probabilities f_t/c_t (as in Equation 6). The transition can be more easily understood when following these five steps (to simplify the notation, we consider a model with only two regression coefficients β_M and β_C).

1. Due to the encompassing prior approach, the prior under M_t is equal to the encompassing prior divided by the prior probability that the order constraints hold c_t—i.e., $\Pr(\beta_M, \beta_C|M_t) = \Pr(\beta_M, \beta_C|M_u)/c_t$ in the constrained area, and zero outside the constrained area. For example, when considering model M_2:$\beta_M > \beta_C$, the prior distribution under M_2 is equal to the encompassing prior divided by $c_2 = 1/2$. Consequently the prior under M_2 is twice as large as the encompassing prior in the constrained region $\beta_M > \beta_C$ and zero on the other half where $\beta_M \geq \beta_C$. This results in a prior for M_2 that adds up to one, which is an essential condition for a probability distribution. Similarly, the posterior under the constrained model M_t is equal to the unconstrained posterior under M_u divided by the posterior probability that the order constraints hold f_t—i.e., $\Pr(\beta_M, \beta_C|\text{data}, M_t) = \Pr(\beta_M, \beta_C|\text{data}, M_u)/f_t$ in the constrained area, and zero elsewhere.
2. The likelihood under the constrained model is equal to the likelihood under the unconstrained model in the constrained area, $\Pr(\text{data}|\beta_M, \beta_C, M_t) = \Pr(\text{data}|\beta_M, \beta_C, M_u)$ if β_M and β_C are in the constrained area of M_t.
3. Bayes' theorem in Equation 4 can be rewritten so that the marginal likelihoods of M_u and M_t equal the prior times the likelihood divided by the posterior (Chib, 1995),

$$\Pr\left(\text{data}|M_u\right) = \frac{\Pr\left(\beta_M, \beta_C|M_u\right) \times \Pr\left(\text{data}|\beta_M, \beta_C, M_u\right)}{\Pr\left(\beta_M, \beta_C|\text{data}, M_u\right)}$$

and

$$\Pr\left(\text{data}|M_t\right) = \frac{\Pr\left(\beta_M, \beta_C|M_t\right) \times \Pr\left(\text{data}|\beta_M, \beta_C, M_t\right)}{\Pr\left(\beta_M, \beta_C|\text{data}, M_t\right)}.$$

4. In this formula, we substitute the alternative expression for the prior and posterior under M_t (Point 1), and we substitute the likelihood under M_u for the likelihood under M_t:

$$\Pr\left(\text{data}|M_t\right) = \frac{\Pr\left(\beta_M, \beta_C|M_u\right)/c_t \times \Pr\left(\text{data}|\beta_M, \beta_C, M_u\right)}{\Pr\left(\beta_M, \beta_C|\text{data}, M_u\right)/f_t}.$$

5. When dividing the marginal likelihood of M_t by the marginal likelihood of M_u, the encompassing prior $\Pr(\beta_M, \beta_C|M_u)$, the unconstrained posterior $\Pr(\beta_M, \beta_C|\text{data}, M_u)$, and the likelihood $\Pr(\text{data}|\beta_M, \beta_C, M_u)$ cancel out, and we obtain the ratio of posterior and prior probability that the order constraints hold under M_u:

$$\text{BF}\left(M_t, M_u\right) = \frac{\Pr\left(\text{data}|M_t\right)}{\Pr\left(\text{data}|M_u\right)} = \frac{f_t}{c_t}.$$

Note

1. For a more elaborate introduction, see also Zyphur and Oswald (2015) in this volume, Kruschke, Aguinis, and Joo (2012), or a classic reference such as Gelman, Carlin, Stern, and Rubin (2004).

References

Adler, P. S., & Borys, B. 1996. Two types of bureaucracy: Enabling and coercive. *Administrative Science Quarterly*, 41: 61-89.

Amstad, F. T., Meier, L. L., Fasel, U., Elfering, A., & Semmer, N. K. 2011. A meta-analysis of work-family conflict and various outcomes with a special emphasis on cross-domain versus matching-domain relations. *Journal of Occupational Health Psychology*, 16: 151-169.

Aquino, K., Tripp, T., & Bies, R. 2001. How employees respond to personal offense: The effects of blame attribution, victim status, and offender status on revenge and reconciliation in the workplace. *Journal of Applied Psychology*, 86: 52-59.

Berger, J. O., & Pericchi, L. 2004. Training samples in objective Bayesian model selection. *The Annals of Statistics*, 32: 841-869.

Bollen, K. 1989. *Structural equations with latent variables*. New York: Wiley.

Bowling, N., & Beehr, T. 2006. Workplace harassment from the victim's perspective: A theoretical model and meta-analysis. *Journal of Applied Psychology*, 91: 998-1012.

Bryson, A., Willman, P., Gomez, R., & Kretschmer, T. 2013. The comparative advantage of non-union voice in Britain, 1980-2004. *Industrial Relations*, 52: 194-220.

Budescu, D. V. 1993. Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, 114: 542-551.

Chib, S. 1995. Marginal likelihood from the Gibbs Output. *Journal of the American Statistical Association*, 90: 1313-1321.

Cohen, J. 1992. A power primer. *Psychological Bulletin*, 112: 155-159.

Deci, E. L., & Ryan, R. M. 1995. Human autonomy: The basis for true self-esteem. In M. Kernis (Ed.), *Efficacy, agency, and self-esteem*: 31-49. New York: Plenum.

Edwards, J. R., & Berry, J. W. 2010. The presence of something or the absence of nothing: Increasing theoretical precision in management research. *Organizational Research Methods*, 13: 668-689.

Eraut, M. 2004. Informal learning in the workplace. *Studies in Continuing Education*, 26: 247-273.

Frone, M. R., Russell, M., & Cooper, M. L. 1992. Antecedents and outcomes of work-family conflict: Testing a model of the work-family interface. *Journal of Applied Psychology*, 77: 65-78.

Gelman, A., Carlin, J. B., Stern, H., & Rubin, D. 2004. *Bayesian data analysis* (2nd ed.). London: Chapman & Hall.

Gelman, A., & Stern, H. 2006. The difference between "significant" and "non-significant" is not itself statistically significant. *The American Statistician*, 60: 328-331.

Gettman, H., & Gelfand, M. 2007. When the customer shouldn't be king: Antecedents and consequences of customer sexual harassment. *Journal of Applied Psychology*, 92: 757-770.

Goldman, B., Gutek, B., Stein, J., & Lewis, K. 2006. Employment discrimination in organizations: Antecedents and consequences. *Journal of Management*, 32: 786-830.

Grandey, A. A., Kern, J. H., & Frone, M. R. 2007. Verbal abuse from outsiders versus insiders: Comparing frequency, impact on emotional exhaustion, and the role of emotional labor. *Journal of Applied Psychology*, 12: 63-79.

Griffin, R. 2011. Seeing the wood for the trees: Workplace learning evaluation. *Journal of European Industrial Training*, 35: 841-850.

Hastie, T., Tibshirani, R., & Friedman, J. 2001. *The elements of statistical learning: Data mining, inference and prediction*. New York: Springer.

Hershcovis, M. S., & Barling, J. 2010. Towards a multi-foci approach to workplace aggression: A meta-analytic review of outcomes from different perpetrators. *Journal of Organizational Behavior*, 31: 24-44.

Heyes, J. 2001. Workplace industrial relations and training. In H. Rainbird (Ed.), *Training in the workplace: 148-168*. London: MacMillan.

Hoijtink, H. 2011. *Informative hypotheses: Theory and practice for behavioral and social scientists*. New York: Chapman & Hall/CRC Press.

Hu, L. T., & Bentler, P. M. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6: 1-55.

Hubbard, R., & Armstrong, J. S. 2006. Why we don't really know what statistical significance means: Implications for educators. *Journal of Marketing Education*, 28: 114-120.

Jeffreys, H. 1961. *Theory of probability* (3rd ed.). New York: Oxford University Press.

Johnson, J. W. 2000. A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, 35: 1-19.

Johnson, J. W., & LeBreton, J. M. 2004. History and use of relative importance indices in organizational research. *Organizational Research Methods*, 7: 238-257.

Johnson, S., Osborn, D. P., Araya, R., Wearn, E., Paul, M., Stafford, M., Wellman, N., Nolan, F., Killaspy, H., Lloyd-Evans, B., Anderson, E., & Wood, S. J. 2012. Morale in the English mental health workforce: Questionnaire survey. *British Journal of Psychiatry*, 201(3): 239-246.

Kammers, M., Mulder, J., de Vignemont, F., & Dijkerman, H. 2009. The weight of representing the body: Addressing the potentially indefinite number of body representations in healthy individuals. *Experimental Brain Research*, 204: 333-342.

Karasek, R. 1979. Job demands, job decision latitude and mental strain: Implications for job redesign. *Administrative Science Quarterly*, 24: 285-308.

Kass, R., & Raftery, A. 1995. Bayes factors. *Journal of the American Statistical Association*, 90: 773-795.

Klugkist, I., Kato, B., & Hoijtink, H. 2005. Bayesian model selection using encompassing priors. *Statistica Neerlandica*, 59: 57-69.

Kluytmans, A., van de Schoot, R., Mulder, J., & Hoijtink, H. 2012. Illustrating Bayesian evaluation of informative hypotheses for regression models. *Frontiers in Psychology*, 3. doi:10.3389/fpsyg.2012.00002

Kruschke, J., Aguinis, H., & Joo, H. 2012. The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods*, 15: 722-752.

Lodewyckx, T., Kim, W., Lee, M. D., Tuerlinckx, F., Kuppens, P., & Wagenmakers, E.- J. 2011. A tutorial on Bayes factor estimation with the product space methods. *Journal of Mathematical Psychology*, 55: 331-347.

Massaro, D. W., Cohen, M. M., Campbell, C. S., & Rodriguez, T. 2001. Bayes factor of model selection validates flmp. *Psychonomic Bulletin & Review*, 8: 1-17.

Mulder, J., Hoijtink, H., & de Leeuw, C. 2012. BIEMS: A Fortran 90 program for calculating Bayes factors for inequality and equality constrained models. *Journal of Statistical Software*, 46:1-39.

Mulder, J., Hoijtink, H., & Klugkist, I. 2010. Equality and inequality constrained multivariate linear models: Objective model selection using constrained posterior priors. *Journal of Statistical Planning and Inference*, 140: 887-906.

Mulder, J., Klugkist, I., van de Schoot, R., Meeus, W., Selfhout, M., & Hoijtink, H. 2009. Bayesian model selection of informative hypotheses for repeated measurements. *Journal of Mathematical Psychology*, 53: 530-546.

Muthén, B., & Asparouhov, T. 2012. Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods*, 17: 313-335.

National Health Service. 2006. *National Health Service national staff survey, 2006*. Available at http://discover.ukdataservice.ac.uk/catalogue/?sn=5736&type=Data%20catalogue [accessed August 30, 2013].

National Institute of Economic and Social Research. 2004. *Workplace employment relations survey 2004* (WERS2004). London: Author. Available at http://www.wers2004.info [accessed August 30, 2013].

Pitt, M. A., & Myung, I. J. 2002. When a good fit can be bad. *Trends in Cognitive Sciences*, 6: 421-425.

Posthuma, R., Campion, M. C., Masimova, M., & Campion, M. A. 2013. A high performance work practices taxonomy: Integrating the literature and directing future research. *Journal of Management*, 39: 1184-1220.

Press, S. J. 2005. *Applied multivariate analysis: Using Bayesian and frequentist methods of inference*. Malabar, FL: Krieger.

Raftery, A. 1993. Bayesian model selection in structural equation models. In K. Bollen & J. S. Long (Eds.), *Testing structural equation models: 163-180*. Beverly Hills, CA: Sage.

Rosenthal, R., Rosnow, R., & Rubin, D. 2000. *Contrasts and effect sizes in behavioral research: A correlational approach*. Cambridge, UK: Cambridge University Press.

Rousseau, D. M. 1995. *Psychological contracts in organizations: Understanding written and unwritten agreements*. Thousand Oaks, CA: Sage.

Russell, J. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39: 1161-1178.

Sellke, P., Bayarri, M., & Berger, J. O. 2001. Calibration of p-values for testing precise null hypotheses. *The American Statistician*, 55: 62-71.

Spiegelhalter, D. J., Abrams, K. R., & Myles, J. P. 2004. *Bayesian approaches to clinical trials and health-care evaluation*. Hoboken, NJ: John Wiley.

Stephens, M., & Balding, D. 2009. Bayesian statistical methods for genetics association studies. *Nature Review Genetics*, 10: 681-690.

Thompson, V. A. 1965. Bureaucracy and innovation. *Administrative Science Quarterly*, 10: 1-20.

Vandenberg, R. J., & Lance, C. E. 2000. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3: 4-70.

Van de Schoot, R., Hoijtink, H., Hallquist, M., & Boelen, P. 2012. Bayesian evaluation of inequality-constrained hypotheses in SEM models. *Structural Equation Modeling: A Multidisciplinary Journal*, 19: 593-609.

Warr, P. 1990. The measurement of well-being and other aspects of mental health. *Journal of Occupational Psychology*, 63: 193-210.

Warr, P. 2007. *Work, happiness, and unhappiness*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wood, S. 2013. HRM, organizational performance and employee involvement. In C. Frege & J. Kelly (Eds.), *Comparative employment relations in the global economy*: 89-107. London: Taylor & Francis.

Wood, S., Braeken, J., & Niven, K. 2013. Discrimination and well-being in organizations: Testing the differential power and organizational justice theories of workplace aggression. *Journal of Business Ethics*, 115: 617-634.

Wood, S., Stride, C., & Johnson, S. 2012. Getting the measure of personal and team morale. *Journal of Health Management*, 14: 535-557.

Zyphur, M. J., & Oswald, F. L. 2015. Bayesian estimation and inference: A user's guide. *Journal of Management*, 41: 390-420. doi:10.1177/0149206313501200