

A Systematic Review of Bayesian Articles in Psychology: The Last 25 Years

Rens van de Schoot
Utrecht University and North-West University

Sonja D. Winter, Oisín Ryan, and
Mariëlle Zondervan-Zwijenburg
Utrecht University

Sarah Depaoli
University of California, Merced

Abstract

Although the statistical tools most often used by researchers in the field of psychology over the last 25 years are based on frequentist statistics, it is often claimed that the alternative Bayesian approach to statistics is gaining in popularity. In the current article, we investigated this claim by performing the very first systematic review of Bayesian psychological articles published between 1990 and 2015 ($n = 1,579$). We aim to provide a thorough presentation of the role Bayesian statistics plays in psychology. This historical assessment allows us to identify trends and see how Bayesian methods have been integrated into psychological research in the context of different statistical frameworks (e.g., hypothesis testing, cognitive models, IRT, SEM, etc.). We also describe take-home messages and provide “big-picture” recommendations to the field as Bayesian statistics becomes more popular. Our review indicated that Bayesian statistics is used in a variety of contexts across subfields of psychology and related disciplines. There are many different reasons why one might choose to use Bayes (e.g., the use of priors, estimating otherwise intractable models, modeling uncertainty, etc.). We found in this review that the use of Bayes has increased and broadened in the sense that this methodology can be used in a flexible manner to tackle many different forms of questions. We hope this presentation opens the door for a larger discussion regarding the current state of Bayesian statistics, as well as future trends.

Translational Abstract

Over 250 years ago, Bayes (or Price, or Laplace) introduced a method to take prior knowledge into account in data analysis. Although these ideas and Bayes’s theorem have been longstanding within the fields of mathematics and statistics, these tools have not been at the forefront of modern-day applied psychological research. It was frequentist statistics (i.e., p values and null hypothesis testing; developed by Fisher, Neyman, and Pearson long after Bayes’s theorem), which has dominated the field of Psychology throughout the 21st century. However, it is often claimed by ‘Bayesians’ that the alternative Bayesian approach to statistics is gaining in popularity. In the current article, we investigated this claim by performing the very first systematic review of Bayesian psychological articles published between 1990 and 2015 ($n = 1,579$). Our findings showed that there was some merit in this thought. In fact, the use of Bayesian methods in applied Psychological work has steadily increased since the nineties and is currently taking flight. It was clear in this review that Bayesian statistics is used in a variety of contexts across subfields of Psychology and related disciplines. This is an exciting time, where we can watch the field of applied statistics change more than ever before. The way in which researchers think about and answer substantive inquiries is slowly taking on a new philosophical meaning that now incorporates previous knowledge and opinions into the estimation process. We hope this presentation opens the door for a larger discussion regarding the current state of Bayesian statistics, as well as future trends.

Keywords: Bayes’s theorem, prior, posterior, MCMC-methods, systematic review

Supplemental materials: <http://dx.doi.org/10.1037/met0000100.supp>

Rens van de Schoot, Department of Methods and Statistics, Utrecht University, and Optentia Research Program, Faculty of Humanities, North-West University; Sonja D. Winter, Oisín Ryan, and Mariëlle Zondervan-Zwijenburg, Department of Methods and Statistics, Utrecht University; Sarah Depaoli, Department of Psychological Sciences, University of California, Merced.

The first author was supported by a grant from the Netherlands organization for scientific research: NWO-VIDI-452-14-006. Preliminary results

were presented during the 7th Mplus Users Meeting organized at Utrecht University, the Netherlands; during the 2016-edition of the European Stats Camp organized by Yhat Enterprises, LLC; and during the 2016-edition of the Mplus Utrecht Summer School.

Correspondence concerning this article should be addressed to Rens van de Schoot, Department of Methods and Statistics, Utrecht University, P.O. Box 80.140, 3508 TC, Utrecht, the Netherlands. E-mail: a.g.j.vandeschoot@uu.nl

. . . whereas the 20th century was dominated by NHST, the 21st century is becoming Bayesian (as forecast by Lindley . . .).
—(Kruschke, 2011, p. 272)

Over 250 years ago, Richard Price published an article written by Thomas Bayes on inverse probability (Bayes & Price, 1763), and just over 200 years ago Simon-Pierre Laplace published the theorem that we now recognize as Bayes's theorem (Laplace, 1814). Bayesian methods implement Bayes's theorem, which states that the data moderate prior beliefs regarding the model parameters, and this process produces updated beliefs about model parameters. The choice of a prior is based on how much information we believe we have preceding data collection, as well as how accurate we believe that information to be. Within Bayesian statistics, priors can come from any source; for example, a meta-analysis, a clinical study or, in the absence of empirical data, expert consensus.

Although the ideas of inverse probability and Bayes's theorem have been longstanding within mathematics, these tools have not been at the forefront of modern-day applied statistics. Applications of Bayesian statistics in psychology date back at least to Edwards, Lindman, and Savage (1963). However, frequentist statistics (i.e., p values and null hypothesis testing; developed by Fisher, Neyman, and Pearson long after Bayes's theorem), have dominated the field of psychology throughout the 21st century. In contrast to the Bayesian paradigm, frequentist statistics associate probability with long-run frequency. The most often used example of long-run frequency is the notion of an infinite coin toss: A sample space of possible outcomes (heads and tails) is enumerated, and the probability of an outcome represents the proportion of the particular outcome divided by the total number of coin tosses. In contrast, the Bayesian paradigm does not carry this notion of long-run fre-

quency. Rather, the Bayesian framework uses prior information and updates this information with new data. For a philosophical discussion on these topics, we refer the interested reader to: Gelman and Shalizi (2013); Haig (2009); Kennedy (2014); McFall and Treat (1999); Morey and Rouder (2011); or Wagenmakers, Lee, Lodewyckx, and Iverson (2008).

A steady increase in the popularity of Bayesian statistics has been noted in systematic reviews conducted outside of psychology. For example, in the field of organizational science, Kruschke (2010) found 42 articles published in 15 different journals between 2001 and 2010 applying Bayesian statistics. Rupp, Dey, and Zumbo (2004) discussed 12 Bayesian applications of IRT models (although these were not necessarily field-specific). Spiegelhalter, Myles, Jones, and Abrams (2000) identified 30 Bayesian application articles published in the field of health technology assessment between 1973 and 1998. Rietbergen (2016) focused on the use and reporting of Bayesian methods in 12 epidemiological and medical journals between 2005 and 2013. She found a total of 86 articles and subsequently reported that several of the articles presented incomplete Method and Results sections. Finally, Ashby (2006) wrote a literature review on Bayesian statistics in medicine spanning from 1982 to 2006. She concluded that Bayesian statistics have pervaded all major areas of medical statistics; including, for example clinical trials, epidemiology, spatial modeling, and molecular genetics. These findings are supported by an initial search on Scopus with the search word "Bayesian" (excluding "Bayesian information criterion"). The results of which are shown in Figure 1; here we can see a steep increase of Bayesian articles over time in many different disciplines.

In the current article, we investigate Kruschke's claim that "the 21st century is becoming Bayesian" by performing a systematic

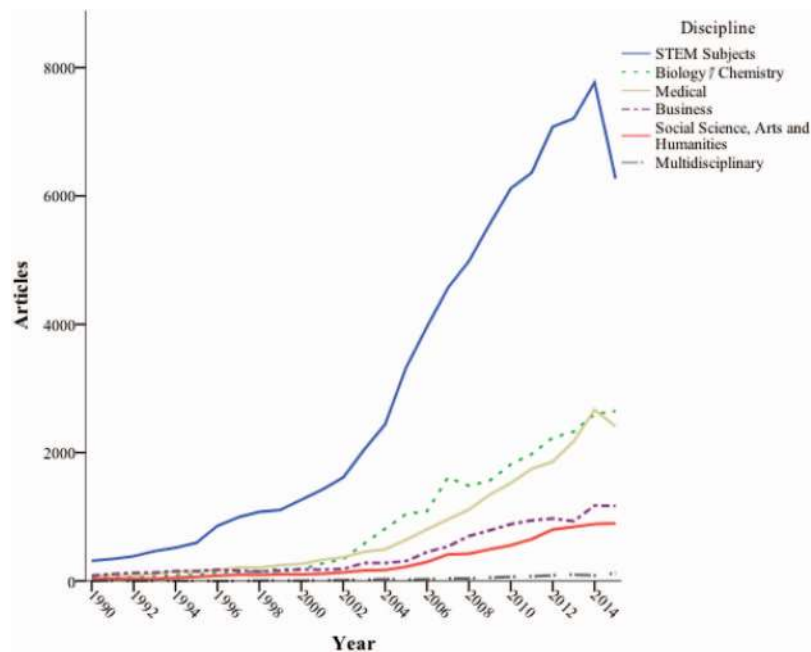


Figure 1. Initial search on Scopus with the search word "Bayesian" in the title, abstract, or keywords (excluding "Bayesian Information Criterion"). STEM = Science, Technology, Engineering, and Mathematics. See the online article for the color version of this figure.

review of Bayesian articles published in the field of psychology. We have two aims with our review: (a) to investigate usage patterns of Bayesian statistics within the field of psychology, and (b) to identify trends over time regarding the use of Bayesian statistics in psychology. To address this latter aim, we detail growth patterns within many different subcategories, including the use of Bayes with different statistical techniques and the use in different subfields of psychology. This aim is informative with regards to the past and current patterns of Bayes usage within psychology, and it is our hope that many different insights can be drawn from this information regarding the presence of Bayesian statistics within the field. We also describe take-home messages and provide “big-picture” recommendations to the field as Bayesian statistics (presumably) becomes more popular within the psychological literature.

Next, we describe the procedure used for our systematic review, followed by the numeric results. Then we elaborate on trends we identified in the Bayesian psychological literature. Based on our findings, we provide some best practices and recommendations for future research in the Discussion section. We provide detailed supplementary materials: (a) A list of all articles we found in the systematic review, including our categorization; (b) A list of journals that have published Bayesian articles; (c) A list of tutorial articles; (d) Examples of empirical articles that can be used as inspiration for how to report Bayesian methods and results, and (e) All of the information needed to reproduce our systematic search.

Method

Step 1: Search Strategy

The search for Bayesian applications was based on the Scopus database of articles published between 1990 and 2015. Articles eligible for inclusion mentioned “Bayesian,” “Gibbs sampler,” “MCMC,” “prior distribution,” or “posterior distribution” in the

title, abstract, or keywords. Note that the MCMC estimation algorithm can be used for Bayesian or frequentist estimation. However, in the current article, we refer to the cases where the MCMC estimation algorithm is implemented in conjunction with observed data, and a prior, in order to sample from the posterior distribution of a particular model parameter. The articles we identified were published in a peer-reviewed journal with “psychology” listed in Scopus as at least one of the journal’s topics; however, the topic of the article could have also included: “arts and humanities,” “business,” “decision sciences,” “economics,” or “sociology.” Articles that mentioned “Bayesian information criterion” as the sole reference to the use of Bayesian statistics were excluded in the search. All steps used for identifying articles are detailed below and in the PRISMA flowchart presented in Figure 2. The exact search terms we used in Scopus can be found in the online supplementary material, including the search used to construct Figure 1 and a file containing references for all of the identified articles.

Step 2: Initial Inclusion and Exclusion Criteria

After all relevant articles were identified, duplicates were excluded, followed by any document-types that were not peer-reviewed articles and were returned by Scopus in error. In addition, the search still extracted some articles solely using the term “Bayesian information criterion” (BIC), despite the exclusion term specified in our search. These articles were screened and excluded from further analysis. The initial screening of the articles was done based on information provided in the abstract. If classification remained unclear after reading the abstract, then we downloaded the full article to further assess classification. Any inaccessible articles were searched for multiple times through various methods (i.e., ResearchGate, Google Scholar). The first authors of the inaccessible articles ($n = 18$) were contacted via a last-known e-mail address. Of those, three articles were found at a later time and 10 were obtained directly from the first author. Thus, in the

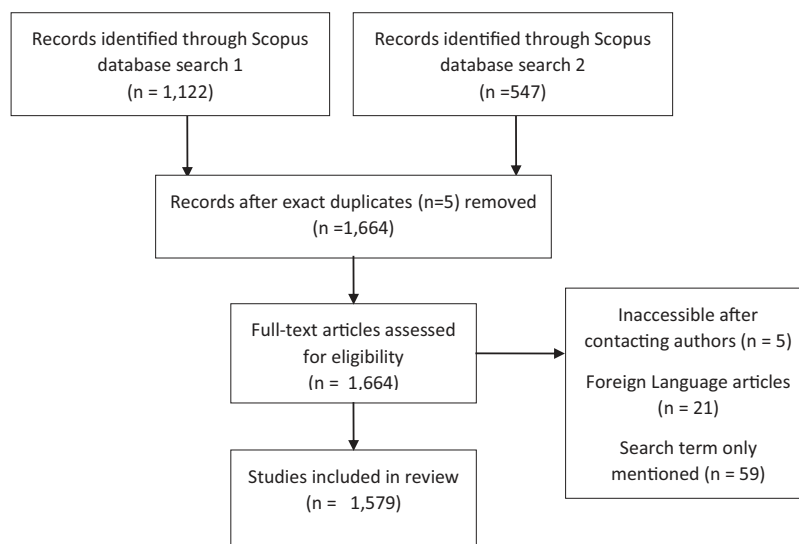


Figure 2. PRISMA flowchart. Search 1 refers to our initial Scopus search (1998–2013 and only the word “Bayesian”). Search 2 was conducted as part of the review process (extending the number of years and the search words). See for more details the online supplementary materials.

end, only five articles remained inaccessible. Moreover, articles not published in English ($n = 21$), or only mentioning Bayes (or any of the other search terms) somewhere in the article but without actually using Bayesian methods ($n = 59$), were excluded. These inclusion criteria left us with 1,579 eligible articles (see Figure 2).

Step 3: Assessment of Bayesian Usage

In the third stage of assessment, the full-text articles were read and screened for Bayesian content by one of the authors, and then double-checked by a second author. We identified several categories representing different types of Bayesian articles.¹ When it was not clear which category an article belonged to, then this was discussed among the authors until consensus was reached. Following the identification and retrieval of eligible articles, we assessed the manner in which Bayesian statistics was used in order to identify articles implementing Bayesian methodology.

Step 4: Detecting Trends

Upon the initial categorization, data extracted from the articles were rereviewed to detect any emerging trends over time.² Even though trends could be roughly identified in the abstracts, it was necessary to read the full text of hundreds of articles in order to adequately describe these trends. Trend information for each article was summarized in Excel files by one of the authors. If something regarding the trend was not clear, then a second author read the article to aid in interpretation and identification of trends. The overall summary written for each article in the Excel files was thoroughly examined by the first author to ensure consistency across all summaries.

Results

The results of this screening process for the 1,579 eligible articles are shown in Table 1. We can see that almost half of the eligible Bayesian articles made use or mention of a regression-based statistical model in relation to Bayesian statistics. Here, we refer to “regression-based” models as the broad class of statistical models, where regression underlies the composition of the model. Some examples of regression-based models in the context of this article are: regression analysis, analysis of variance (ANOVA), confirmatory factor analysis (CFA), structural equation modeling (SEM), item response theory (IRT), and multilevel modeling. All of these articles used MCMC techniques (e.g., Gibbs sampling or

Metropolis Hastings) to estimate model parameters. These types of regression-based models differ from, for example, machine learning techniques, which may not contain regression elements in the statistical process.

Of these regression-based articles, we identified four distinct categories. The breakdown of each of these article types is shown in Table 2. “Empirical” articles were defined as those utilizing Bayesian methods in the analysis of real data in order to answer a substantive research question. “Tutorial” articles (see the online supplementary material for a full list) were defined as step-by-step explanations of the use of Bayesian statistics meant for new users of the method. Note that we do not claim to have included all possible tutorials published about Bayesian analyses because many of these have presumably been published outside of the field of psychology. “Simulation” articles were defined as simulation studies introducing and assessing a new technique using Bayesian MCMC methods. “Theoretical/technical” articles were defined as those only using formulae or mathematical proofs, with more technical details than tutorial articles and with a target audience of methodologists/statisticians. In the Trends section we discuss the articles in these categories in great detail. While reading the regression-based Bayesian articles, we also found some secondary categories, which include “Bayesian Meta-Analysis” and “Commentary” articles, but because these are only small categories we refrain from discussing them in detail. However, we do include them in Table 2 for reference.

We can see from Table 1 that there are two other sizable categories which we identified: “Bayes as cognition/learning” and “Computational model.” The former refers to articles that discuss Bayesian statistics, but only as a model explaining human perception or reasoning. In addition, the latter category contains articles using Bayesian statistics to explain how humans reason (see, e.g., Albert, 2000), or articles modeling cognition using a Bayesian computation model—that is, a task was set and a model was defined, which attempted to imitate a human “Bayesian” thinker in overcoming or assessing the task (see, e.g., Kemp & Tenenbaum, 2009). These should be seen as categories that apply Bayesian methods in cognitive psychology, even if the application is made in slightly different ways. In the Trends section, we discuss these two categories in great detail.

Furthermore, 77 articles were concerned with Bayes Networks analysis, which are part of the family of probabilistic graphical models (PGMs). Each node in a Bayesian network represents a random variable, while each connecting line represents a probabilistic dependency between random variables. A PGM is called a Bayes Net when the graph is a directed acyclic graph (DAG), that is, a graph where it is impossible to form a cycle between random variables. Bayes nets are applied in many fields, such as education (Almond, DiBello, Moulder, & Zapata-Rivera, 2007), social psy-

Table 1
Breakdown of Bayesian Usage in 1,579 Eligible Articles in Subcategories

Use of Bayesian statistics	<i>N</i>	%
Regression-based	740	46.9
Bayesian methods in cognitive psychology		
Bayes as cognition/learning	456	28.9
Computational model	175	11.1
Bayesian network	77	4.9
Direct application of Bayes theorem	54	3.4
Speech/image recognition	46	2.9
Machine learning	18	1.1
Bayesian model averaging	13	.8

¹ When we started the review we first screened the articles published in 1998, 2003, 2008, and 2013 and we determined categories during this phase. Next, these categories were applied toward the rest of the review, and new categories were added as we saw new developments emerge in the field. For example, the “theoretical/simulation/empirical” classification was made a priori, the only change being meta-analysis which was taken out of the empirical category and made its own category, commentary articles were initially classified tutorial articles, and the human/nonhuman subclassification was added later.

² We thank the editor for helping us detect trends.

Table 2
Breakdown of Article Types in 740 Eligible
Regression-Based Articles

Use of Bayesian statistics	<i>N</i>	%
Empirical		
Sample is human	167	22.6
Sample is nonhuman	37	5.0
Tutorial	100	14.4
Simulation	198	25.9
Technical/theoretical	208	27.8
Meta-analysis	13	1.8
Commentary	17	2.6

chology (Krivitsky, Handcock, Raftery, & Hoff, 2009), cognitive psychology (Fenton, Neil, & Lagnado, 2013), and neuroscience (Zhang et al., 2015). Even though the term “Bayesian Network” was first used by Judea Pearl in 1985 (Pearl, 1985), and we were able to find applications as early as 1993 (Eizirik, Barbosa, & Mendes, 1993), our review shows that it is only after the turn of the 21st century that Bayesian Networks started to play a role in the field of psychology.

Finally, there are some smaller categories we identified during our search. One such category contains articles that used Bayesian models in relation to speech or image recognition. These articles included theoretical models of human speech or image recognition, as well as computational models for image and speech recognition (e.g., Yu & Huang, 2003). Another small category deals solely with Bayes Factors, which are calculated from the ratio of BICs for two models, as described by Kass and Raftery (1995; see, for example, Cipora & Nuerk, 2013). Other small categories contain articles in the areas of Bayesian model averaging (e.g., Wasserman, 2000), or machine learning (e.g., Garrard, Rentoumi, Geserich, Miller, & Gorno-Tempini, 2014). Finally, there is a category of articles where Bayes’ formula is directly applied by analytically deriving the posterior through “simply” filling in Bayes’s theorem with exact values; some articles also calculated the likelihood without sampling. Such articles are labeled in Table 1 as “Direct application of Bayes theorem.” Examples of these applications directly solving Bayes’s theorem include Allen and Iacono (1997) and Morrell, Taylor, Quine, Kerr, and Western (1994). More recently, Bayes’ formula has been applied as a way to compute standard measures of sensitivity, specificity, posterior predictive values, and accurate differentiation of a certain measure; some examples include Tondo, Visioli, Preti, and Baldessarini (2014) and Serra et al. (2015). In the earlier articles in our sample, the use of direct application can be explained by the fact that there were not many statistical tools available for assessing data through Bayesian methods (i.e., the Gibbs sampler and other similar sampling methods were not yet available). Because these smaller categories contained too few articles to gather trends from, we refrain from describing them any further.

Trends

This section shows a detailed breakdown of the role Bayes has played in different areas of statistics implemented within psychology (as defined by Scopus). Throughout the review process, we have identified several different categories of published Bayesian

articles. Within and across these categories, there were many interesting trends. Within these trends, we mainly focused on how results from technical/simulation articles started to influence applied articles. We present these trends next, and every subsection also describes take-home messages and provides “big-picture” recommendations as Bayesian statistics (presumably) becomes more popular within psychology (see italicized text at the end of each subsection for these messages).

The Use of Bayesian Estimation Over the Years

As can be seen in Figure 3, not only has the absolute number of articles using Bayesian statistics increased (Figure 3A), but the proportion of Bayesian articles relative to the total number of articles has also increased (Figure 3B).³ Figure 3C presents all articles included in our review split over nine categories, which range from regression-based articles to machine learning. Figure 3D focuses our attention specifically on regression-based articles. Within this category, we see that after a slow start, empirical articles have taken flight and are currently the most frequently published article type. Theoretical and simulation articles were relatively more prevalent earlier in this field, but these types of articles do not show growth quite as explosively compared to empirical articles. Notably, the first empirical applications (according to our definition) were not published until the year 2000 (Smith, Kohn, & Mathur, 2000; Verguts & De Boeck, 2000).⁴ Each of these earliest examples used Bayesian statistics as a tool to solve computation problems with complex models.

In conclusion, the use of Bayesian methods is indeed increasing in absolute and relative numbers, and empirical regression-based applications are especially taking flight.

“Bayesian” Journals

The 1,579 articles identified in our systematic review were published in 269 different journals. However, 50% of all articles were published in just 19 of these journals. A total overview of journals that have published Bayesian work can be found in the online material; this is a good resource for the types of journals open to publishing applied Bayesian inquiries. In what follows, we describe several additional trends found in our review.

JCR Subject Categories

In Table 3 we list journals with 10 or more Bayesian publications since 1990 ($n = 37$) in order to exclude journals that only sporadically publish a Bayesian article. We extracted subject category, impact factor, and ranking from the JCR Social Science Edition, 2014 database. As Table 3 shows, a great majority of the

³ To create these figures, we first searched Scopus for the word “Bayesian” and only included psychology articles. Next, we searched Scopus for an empty string, again only including psychology articles, to approximate the total number of articles published in psychology per year. For more information, please refer to the online supplementary materials.

⁴ Although it could be argued that articles such as Hoijtink and Molenaar (1997) or Arminger and Muthén (1998) are empirical articles, we only defined an article as “empirical” if the main research question was a substantive question and the introduction section was focused on the underlying substantive theories being examined.

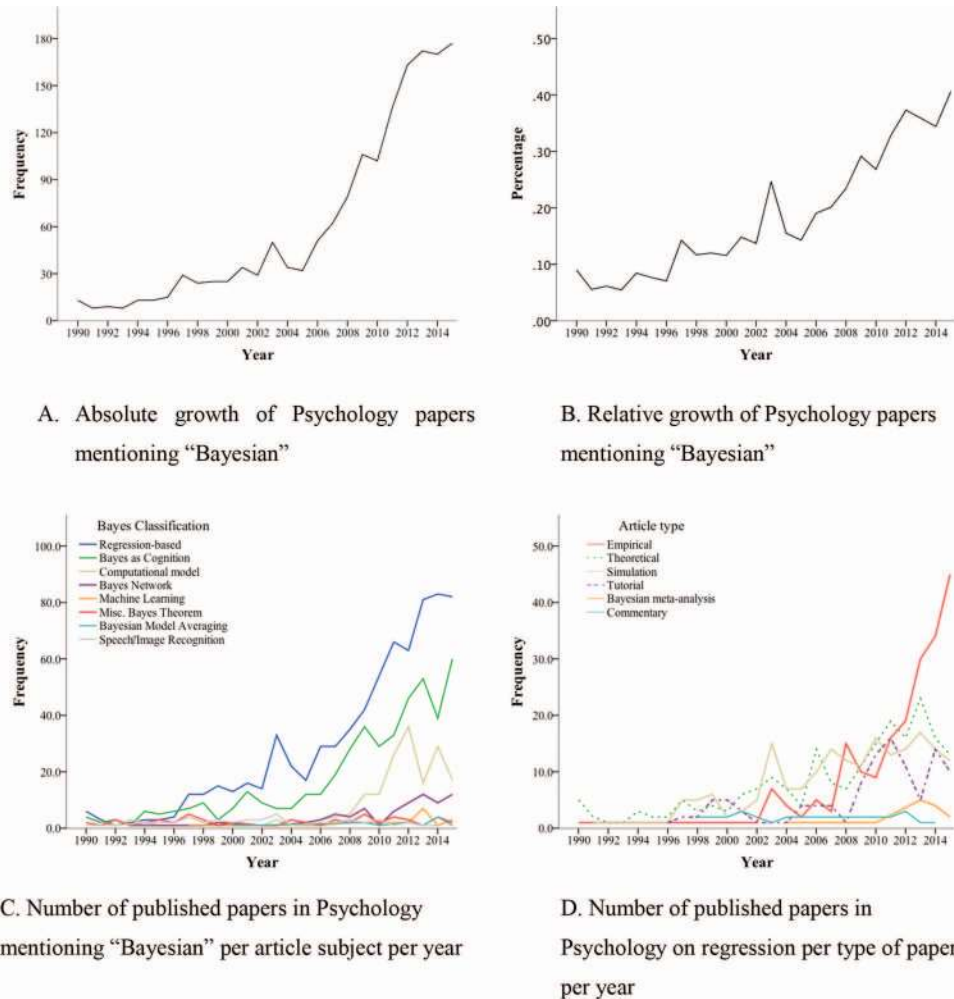


Figure 3. Evolution of articles using Bayesian statistics in the field of psychology (A and B) divided over subject category (C) and, for regression-based, article type (D). See the online article for the color version of this figure.

37 journals ranked in the top 50% ($n = 31$) of journals of at least one of their assigned subject categories; 19 fell in the top 25%.⁵ In the category “mathematical psychology,” three of the top four journals regularly publish Bayesian articles, namely: *Psychonomic Bulletin and Review* (1st), *Behavior Research Methods* (2nd), and *British Journal of Mathematical and Statistical Psychology* (4th).

Table 4 presents an overview of the JCR subject categories related to the subfield of psychology where the articles reported in Table 3 were published. Based on the absolute number of journals, “experimental psychology” is the most popular subject category, where 18 out of 85 journals regularly publish Bayesian articles. However, based on the percentage of journals within a category, “mathematical psychology” has the highest percentage (38.46%) of journals in this subject area publishing Bayesian articles. If we focus exclusively on journals categorized within psychology, areas less influenced by Bayesian methods are: “biological,” “developmental,” “general,” and “social psychology.” Two JCR categories were completely absent: “psychology, clinical” and “psychology, psychoanalysis.” However, our search showed that there are many

more journals with less than 10 articles published (232 journals), which could potentially fall within these subfields of psychology. As we will see later when we discuss the topic areas of empirical regression-based articles, there are many developmental and educational articles published. However, these are not often published within the specialized journals in the respective categories, or they are published in journals that are still new to accepting Bayesian publications (i.e., those that have published fewer than 10 Bayesian articles).

Based on our results, it seems that there are journals within most areas of psychology that are open to publishing studies based on Bayesian analysis, and many of these journals rank high within their respective subfields.

⁵ It is interesting to note that JCR categorizes some of these journals outside of the field of psychology, and sometimes even outside of the social sciences. This is at odds with Scopus, which categorized all of our included articles as “psychology”.

Table 3

Journals That Published at Least 10 Bayesian Publications in the Field of Psychology Since 1990: 1,056 Articles out of a Total of 1,579 Articles, and 37 Journals Out of Total of 269 Journals

Journal	Frequency	Category	Ranking in category	Impact factor ¹	Max impact factor ¹	Median impact factor ¹
1. <i>Psychometrika</i>	111	Social sciences, mathematical methods	22/46	1.085	4.176	1.021
2. <i>Journal of Mathematical Psychology</i>	70	Social sciences, mathematical methods	4/46	2.609	4.176	1.021
3. <i>Applied Psychological Measurement</i>	68	Social sciences, mathematical methods	20/46	1.178	4.176	1.021
4. <i>Frontiers in Psychology</i>	63	Psychology, multidisciplinary	23/129	2.56	21.81	1.015
5. <i>Psychological Review</i>	57	Psychology, multidisciplinary	5/129	7.972	21.81	1.015
6. <i>Cognitive Science</i>	55	Psychology, experimental	30/85	2.446	21.965	2.009
7. <i>Cognition</i>	54	Psychology, experimental	9/85	3.479	21.965	2.009
8. <i>Psychonomic Bulletin and Review</i>	36	Psychology, mathematical	1/13	3.369	3.369	1.178
		Psychology, experimental	12/85	3.369	21.965	2.009
9. <i>Frontiers in Human Neuroscience</i>	33	Neurosciences ²	85/252	3.626	31.427	2.791
		Psychology ²	13/76	3.626	21.81	2.03
10. <i>British Journal of Mathematical and Statistical Psychology</i>	32	Psychology, mathematical	4/13	2.167	3.369	1.178
		Psychology, experimental	38/85	2.167	21.965	2.009
11. <i>Psychological Methods</i>	30	Psychology, multidisciplinary	6/129	7.338	21.81	1.015
12. <i>Educational and Psychological Measurement</i>	28	Psychology, educational	30/55	1.154	4.061	1.308
13. <i>Cognitive Processing</i>	27	Psychology, experimental	67/85	1.388	21.965	2.009
14. <i>Speech Communication</i>	25	Acoustics ²	12/31	1.256	4.924	.912
		Computer science, Interdisciplinary applications ²	62/102	1.256	4.925	1.401
15. <i>Journal of Experimental Psychology: Learning Memory and Cognition</i>	24	Psychology, experimental	23/85	2.862	21.965	2.009
16. <i>Topics in Cognitive Science</i>	24	Psychology, experimental	16/85	3.063	21.965	2.009
17. <i>Multivariate Behavioral Research</i>	22	Social sciences, mathematical methods	6/46	2.477	4.176	1.021
		Psychology, experimental	27/85	2.477	21.965	2.009
18. <i>Journal of Experimental Psychology: General</i>	21	Psychology, experimental	3/85	5.929	21.965	2.009
19. <i>Behavior Research Methods</i>	20	Psychology, mathematical	2/13	2.928	3.369	1.178
		Psychology, experimental	20/85	2.928	21.965	2.009
20. <i>Theory and Decision</i>	20	Economics	192/333	.72	6.654	.86
		Social Sciences, Mathematical Methods	36/46	.72	4.176	1.021
21. <i>Decision Support Systems</i>	19	Computer science, artificial intelligence ²	27/123	2.313	8.746	1.406
		Computer science, information systems ²	16/139	2.313	6.806	.971
		Operations research & management science ²	11/81	2.313	4.376	1.079
22. <i>Journal of Educational Measurement</i>	19	Psychology, Educational	36/55	.922	4.061	1.308
		Psychology, Applied	44/76	.922	6.071	1.205
		Psychology, Mathematical	11/13	.922	3.369	1.178
23. <i>Behavioral and Brain Sciences</i>	18	Psychology, Biological	1/14	20.771	20.771	1.917
24. <i>Journal of Classification</i>	18	Psychology, Mathematical	12/13	.727	3.369	1.178
25. <i>Computer Speech and Language</i>	16	Computer science, artificial intelligence ²	47/123	1.753	8.746	1.406
26. <i>Thinking and Reasoning</i>	15	Psychology, Experimental	34/85	2.2	21.965	2.009
27. <i>Technological Forecasting and Social Change</i>	14	Business	31/115	2.058	7.475	1.4
	13	Psychology, Applied	13/76	2.201	6.071	1.205
		Psychology, Social Management	15/62	2.201	6.692	1.522
28. <i>Organizational Behavior and Human Decision Processes</i>			38/185	2.201	7.769	1.208
29. <i>Psychological Science</i>	13	Psychology, multidisciplinary	11/129	4.94	21.81	1.015
30. <i>Acta Psychologica</i>	12	Psychology, experimental	31/85	2.248	21.965	2.009
31. <i>Perception</i>	12	Psychology, experimental	74/85	.906	21.965	2.009
32. <i>Psychological Bulletin</i>	12	Psychology, multidisciplinary	3/129	14.756	21.81	1.015
33. <i>Quarterly Journal of Experimental Psychology</i>	12	Psychology, experimental	36/85	2.127	21.965	2.009
34. <i>Developmental Science</i>	11	Psychology, developmental	7/68	3.808	7.26	1.728
		Psychology, experimental	8/85	3.808	21.965	2.009

(table continues)

Table 3 (continued)

Journal	Frequency	Category	Ranking in category	Impact factor ¹	Max impact factor ¹	Median impact factor ¹
35. <i>Memory and Cognition</i>	11	Psychology, experimental	29/85	2.457	21.965	2.009
36. <i>Social Science and Medicine</i>	11	Social sciences, biomedical	4/39	2.89	5.288	1.311
		Public, environmental and occupational health	16/147	2.89	10.042	1.372
37. <i>Social Networks</i>	10	Anthropology	14/84	2	4.553	.69
		Sociology	8/142	2	4.39	.783

Note. Some journals carry more than one subject category.

¹ All Impact Factors based on 2014. Max and median impact factors are all within their respective category. ² These journals were not part of the JCR Social Science Edition, 2014, but were instead included in the JCR Science Edition, 2014.

Psychometrika and Open Access Journals

Another trend became apparent as we inspected the journals by decade. In the 1990s, Bayesian articles were published in 64 different journals, with *Psychometrika* as the major contributor (17.8%). In the 2000s, the dominance of *Psychometrika* decreased to 9.2% (although still first place), and a wider variety of journals ($n = 139$) published Bayesian articles. In the last 5 years (2011–2015), 191 different journals published Bayesian articles. In addition, there was a shift in the journal-type that published the most Bayesian work during this time, with the top five journals (in order): *Frontiers in Psychology* (6.8%), *Journal of Mathematical Psychology* (4.8%), *Cognition* (4.1%), *Psychometrika* (4%), and *Frontiers in Human Neuroscience* (3.5%).

Based on our review, it appears that the landscape of journals open for Bayesian articles is changing, and that more substantive and open access journals (i.e., the *Frontiers* journals) are becoming popular Bayesian outlets.

Zooming in on Psychological Methods

An important journal for Bayesian publications is *Psychological Methods*, because this journal has broad appeal (and impact) to methodologists, as well as applied psychological researchers. We also highlight this journal because it is an important contributor to Bayesian statistics, especially with the current special issue dedicated solely to this topic.

We present publication rates and citation counts from 1996–2015 for this journal in Table 5 (the Journal was founded in 1996). There appears to be a steady increase in the number of Bayesian

articles (according to Scopus) being published in *Psychological Methods*. From 1996–2008, only 5 of the 13 years had Bayesian articles published. However, from 2009–2015, every year has seen at least one Bayesian article published. This suggests that Bayesian statistics is starting to have a more consistent presence in *Psychological Methods*.

It is not surprising to see that older articles have received more citations compared with those only more recently published. For example, articles published in 1996–2005 tend to have higher citation counts compared to articles published in the last 5 years. Note that the current ISI Impact Factor for *Psychological Methods* is 7.338. We used the JCR database to retrieve the total number of citations for articles published in *Psychological Methods*. This information was available for the years 2009–2013. We used Web of Science, a search engine from the same company as JCR, to extract the number of articles published per year and the number of citations for articles included in our review. Table 5 shows that 4 of the 5 years for which comparisons were possible, Bayesian articles received more citations than the journal average. The exception is 2013, where the journal average is 4.27, but Bayesian articles only received 3.86 citations on average.

Aside from citation and publication rates, we also want to note the types of articles being published specifically in *Psychological Methods*. Of the 32 Bayesian articles published since 2000, 20 (or 62.5%) of these articles included a simulation component. The remaining 12 articles introduced new methods and often included an example component to illustrate the method(s) being introduced. There was no trend of time relating to the type of article published. In other words, simulation studies were not more or less

Table 4
Overview of Distribution of Journals Over ISI Web of Knowledge Categories

Category	Total # of journals with ≥ 10 publications	Total # of journals in category	Percentage
Psychology, experimental	16	85	18.82%
Psychology, mathematical	5	13	38.46%
Psychology, multidisciplinary	5	129	3.88%
Social sciences, mathematical methods	5	46	10.87%
Psychology, educational	2	55	3.64%
Psychology applied	2	76	2.63%
Psychology, biological	1	14	7.14%
Psychology, social	1	62	1.61%
Psychology, developmental	1	68	1.47%
Psychology (General science database)	1	76	1.32%

Table 5
Publication and Citation Information for Bayesian Articles Published in Psychological Methods (1996–2015)

Year	Total # articles appearing in <i>Psychological Methods</i>	Total # Bayesian articles in <i>Psychological Methods</i>	Total % of all articles that were Bayesian in <i>Psychological Methods</i>	Total # of citations per year ^a	Mean # of citations Bayesian articles	Median # of citations Bayesian articles	Total # citations journal JCR	Mean # citations JCR
1996	29	0	0					
1997	27	0	0					
1998	31	0	0					
1999	27	0	0					
2000	27	1	3.7	43	43.00	43.00		
2001	27	0	0					
2002	29	2	6.9	3,560	1,780.00	1,780.00		
2003	36	0	0					
2004	26	1	3.85	23	23.00	23.00		
2005	30	1	3.33	54	54.00	54.00		
2006	31	0	0					
2007	28	1	3.57	5	5.00	5.00		
2008	21	0	0					
2009	24	2	8.33	85	42.50	42.50	321	13.38
2010	30	4	13.33	53	13.25	12.50	286	9.53
2011	31	3	9.67	37	12.33	8.00	259	8.35
2012	45	3	6.67	118	39.33	20.00	358	7.96
2013	33	7	21.21	27	3.86	4.00	141	4.27
2014	35	4	11.43	24	6.00	.00		
2015	32	1	3.13	0	.00	.00		

^aCitation counts from Web of Science as of January 6, 2015.

frequently published in the earlier years compared with the more recent years.

Overall, *Psychological Methods* tends to publish a relatively high proportion of articles that are not based on Bayesian statistics. However, we can see that the proportion of Bayesian articles (mainly simulation-based) may be on the rise and carry a larger impact compared with non-Bayesian articles.

Trends Within Regression-Based Models

Because regression-based articles represent the largest statistical category of our search (see Table 1), we devote an entire section to these articles.

Technical/Theoretical Articles

Any time a new statistical tool or method finds its way into a field, there are typically many technical or theoretical articles written about that tool. In a subsequent section, we also describe the transfer of knowledge obtained from the technical articles to simulation articles (where the new method is tested against alternative methods), as well as to the “real” applications. Here, we highlight the main themes of the technical articles on Bayesian statistics within psychology. We identified 172 articles introducing Bayesian statistics as a new method of estimation. Of these, 29.7% related to item response theory (IRT) models, 17.4% related to structural equation models (SEMs), 13.4% related to multilevel models, 8.7% related to computerized adaptive testing (CAT), and the remaining articles related to a variety of other models.

Overall, we can see that there has been a relatively large number of technical or theoretical articles written within psychol-

ogy that consider Bayesian methods, especially related to some more advanced modeling frameworks such as IRT or SEM.

Simulation Articles

The majority (60.6%; total $n = 198$) of simulation articles did not compare Bayesian estimation against other estimators. Rather, they investigated the performance of a newly proposed Bayesian method for different types of populations (e.g., de Leeuw & Klugkist, 2012), compared different MCMC algorithms with each other (e.g., Arminger & Muthén, 1998), examined different prior specifications (e.g., Lee & Zhu, 2000), or explored different levels of missingness/missing data mechanisms (e.g., Song & Lee, 2002). We found interesting trends within these articles regarding the type of models being examined. Specifically, 66 articles dealt with IRT models, 51 examined SEMs (including mixture, nonlinear, multi-level SEMs, and missing data issues in SEM), and 25 focused specifically on CAT; the remaining articles varied widely in model-type and focus.

We also found 78 articles (39.4%) that compared the performance of Bayesian methods with other estimation methods—almost always maximum likelihood estimation (ML). Based on the abstracts of these articles, 70.5% concluded that Bayesian methods outperformed the comparison method(s). In the remaining abstracts, it was concluded that the performance of Bayesian methods was equivalent to the comparison method (14.1%), or it was stated that it depended on the specific situation (e.g., depending on the prior specification; 7.7%). In six abstracts, it was concluded that the Bayesian method performed worse than the comparison method (7.7%).

In general, Bayesian methods were found to outperform other estimation methods for many different performance criteria that is,

Type I error rates, power, and producing stable and accurate coverage rates) across a wide range of statistical models (e.g., regression, CAT, IRT, SEM). These findings were especially relevant when the sample size was small (we discuss the issue of small samples in a separate subsection), when the model was more complex (e.g., Wang & Nydick, 2015), or in the situation where alternative methods were simply not developed yet (e.g., Wollack, Bolt, Cohen, & Lee, 2002).

Overall, simulation studies are a large part of the regression-based Bayesian articles, and it is shown that Bayesian estimation typically outperforms other estimation methods. However, this is not always the case so researchers should be careful when using Bayesian methods as a “golden” solution to all modeling issues.

Empirical Regression-Type Articles

In this category, we found 167 empirical articles using MCMC techniques (e.g., Gibbs sampling or Metropolis Hastings) to estimate model parameters instead of traditional frequentist methods—that is, in a sense “standard analyses gone Bayesian.”⁶ We discuss several aspects of this large group of articles next.

Field. We clustered the articles in topic-fields contributing to empirical regression-based work. In ranked order, the fields with Bayesian applied articles in this category were: cognitive psychology (24.6%), health psychology (12.0%), developmental psychology (10.2%), educational psychology (6.6%), personality psychology (5.4%), neuropsychology (4.2%), and a variety of 26 smaller fields. The discrepancy between these percentages and our previous discussion on the representation of Bayesian methods across journals is due to some of these articles being published in journals that are not traditionally thought of as “developmental” or “educational.” Rather, they were published in general-topic journals. An example is Knops, Zitzmann, and McCrink (2013), which focuses on child development but was published in *Frontiers in Psychology*. An additional explanation is that these articles were published in journals that have not yet published 10 or more Bayesian articles in the past 25 years, which was one of our criteria for journal discussion. Examples include: *Developmental Psychology* (five publications, where one was regression-based empirical), the *Journal of Educational Psychology* (two publications, both regression-based empirical), and the *Journal of Psychoeducational Assessment* (three publications, all regression-based empirical).

In conclusion, we found applications in many different subfields within psychology, which indicates that Bayesian analysis is slowly becoming accepted across the entire field of psychology.

Reasons why researchers use Bayesian statistics. There are many different reasons provided in the regression-based articles why Bayesian techniques were used. We have sorted through the empirical regression-based articles and have identified the main reasons why Bayesian methods were used. Note that in many articles various reasons were provided and as such the percentages reported in this section do not add up to 100%. In 11.4% of all the article, there was not a clear argument given as to why Bayesian methods were implemented, or it was just mentioned that previous literature advised to use Bayes (4.8%).

The first category of reasons for using Bayesian methods is that researchers may be “forced into” the implementation because some complex models simply cannot be estimated using other

approaches (as is argued in, e.g., Heeren, Maurage, & Philippot, 2013), or is difficult because of computational problems with ML or WLS estimation (e.g., Ryoo et al., 2015). Also, computational burden is mentioned as argument why to use Bayesian estimation (e.g., Choi, Koh, & Lee, 2008), or researchers faced convergence issues with maximum likelihood estimation (e.g., Lee, Choi, & Cho, 2011). At times, models were intractable or not able to be estimated because of the high dimensional numerical integration needed for maximum likelihood estimation (e.g., Humphries, Bruno, Karpievitch, & Wotherspoon, 2015). Bayesian methods were also used to produce more accurate parameter estimates compared to conventional methods (e.g., Wilson, Barrineau, Butner, & Berg, 2014), or to get around issues of model identification (e.g., Wilson et al., 2014). These kinds of arguments are mentioned (as at least one of the reasons) in 27.5% of the empirical regression-based articles we found in our review. We also found several articles where the reason to use Bayesian statistics was because of: violating assumptions in other estimation methods (e.g., Desmet & Feinberg, 2003; 7.8%), the great modeling flexibility of Bayes (e.g., Jackson, Gucciardi, & Dimmock, 2014; 10.8%), missing data handling (e.g., Bulbulia et al., 2013; 3.6%), or improved performance for small sample sizes. Specifically, small samples were mentioned in 14.4% of the empirical articles and we expand on this further in a separate section below where we discuss trends across different types of articles (i.e., theoretical, simulation, and empirical articles).

Another category of reasons reported in the regression-based articles to use Bayesian estimation is because it is appealing to incorporate knowledge into the estimation process via priors. The use of (informative) priors was explicitly mentioned as at least one of the reasons for selecting Bayes in 12.6% of the articles; this included the use of subjective priors (e.g., Cavanagh, Wiecki, Kochar, & Frank, 2014), small-variance priors on cross loadings (e.g., Golay, Reverte, Rossier, Favez, & Lecerf, 2013), or to test for approximate measurement invariance through the use of approximate-zero priors used to deal with noninvariance across groups (e.g., Bujacz, Vittersø, Huta, & Kaczmarek, 2014).

Researchers may also implement Bayesian methods because of the model selection tools available within the estimation framework, such as the DIC or the Bayes Factor. Model selection was reported as at least one of the reasons in 8.4% of the articles, and the Bayes Factors in 15.6%. Another 6% of the articles reported to use Bayes Factors as a way to find evidence in support for the null hypothesis (e.g., van Ravenzwaaij, Dutilh, & Wagenmakers, 2011) and 5.4% of the articles tested informative hypothesis (e.g., Wong & van de Schoot, 2012). We come back to the use of model selection by means of the DIC and Bayes Factors in a separate section.

A third category of reasons to use Bayes is because of modeling uncertainty as part of the statistical model. Eleven articles (6.6%) explicitly reported using Bayesian statistics to deal with uncer-

⁶ In this category we also found 37 nonhuman empirical articles (e.g., gene data, fMRI data, financial data, etc.), and we decided not to discuss these articles in detail here because: (a) our main interest was related to the psychological literature surrounding the study of humans, and (2) we cannot guarantee that our search included all articles published in these areas because the search terms were not specifically geared toward non-human research.

tainty (e.g., Foreman, Morton, & Ford, 2009), or because the posterior has desirable properties to deal with random variables (9%). And another 7.8% of the articles mentioned using Bayesian methods because of the advantage of credibility intervals over confidence intervals (e.g., Lynch, Brown, & Harmsen, 2003). Related to this, the region of practical equivalence (as argued by, e.g., Kruschke, 2011) has also been slowly gaining popularity in the literature (see, e.g., Ahn et al., 2014; Liddell & Kruschke, 2014). In 12.6% of the articles the more intuitive interpretation of Bayesian results was given as the main argument to use Bayesian estimation.

In conclusion, a wide range of reasons is provided to use Bayesian statistics and, moreover, we believe that as Bayesian analyses become more and more accepted as a standard estimation tool (which we feel they are), researchers will feel less pressure to justify why Bayesian methods were used.

Bayesian software being implemented. While the Bayesian statistical program BUGS was already created in 1989,⁷ the more familiar (and most popular) version WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000) was not developed until 1997. Up until 2012, WinBUGS was by far the most popular Bayesian program being used in regression-based articles; it was implemented in almost half of all empirical articles up until this point. During this same time, there was a wide range of software packages and programs that were cited only once or twice, making BUGS the clear favorite, see also the software overview presented in Spiegelhalter et al. (2000). The popularity of WinBUGS has faded post-2012 with the advent and publicizing of alternative programs. From 2013–2014, only 8.8% of empirical articles used WinBUGS, and only 10.0% used it in 2015. Other packages such as JAGS (Plummer, 2016; 12.3% in 2013–14 and 17.5% in 2015) and Mplus (Muthén & Muthén, 1998–2015, 22.8% in 2013–2014 and 20.0% in 2015) are becoming much more popular in the empirical Bayesian literature, with Mplus taking over the leading position from 2013 onward. We also feel that JASP (Love et al., 2015), Stan (Carpenter et al., in press; Kruschke, 2014), and blavaan (Merkle & Rosseel, 2015) are promising new programs that will gain usage over the coming years. It was perhaps most striking in our assessment of Bayesian program usage that 22.2% of the empirical articles we identified did not report software at all.

Overall, the landscape of Bayesian software has changed drastically over the last few years, with more user-friendly programs now available.

Priors. There were a variety of types of priors used in the empirical articles. Notably, 31.1% of the articles did not even discuss the priors implemented. For some of these articles (8.4% of the total amount), we assume from the text that the reason was that default priors from the software were implemented. For example, in software like Mplus and Amos, Bayesian estimation can be requested without having to manually specify prior distributions. This feature makes it really easy to use Bayesian methods in these programs, but many researchers may be unaware that a prior is set for every parameter by the software and these settings might not always be valid (see van de Schoot, Broere, Perryck, Zondervan-Zwijnenburg, & van Loey, 2015 for a detailed discussion). Another 24% of the articles discussed the prior superficially, but did not provide enough information to reproduce the prior settings (i.e., hyperparameters).

Out of all articles, 45% reported details surrounding the priors, including hyperparameters. In half of these articles, all of the prior-related equations were provided. When informative priors were used, see also next paragraph, 73.3% of the articles reported the hyperparameter values for each prior. Overall, only 43.1% of empirical articles reported hyperparameter values. It appeared that in (at least) 37.7% of these articles, a normally distributed prior was used and 7.2% used a uniform prior. For the rest of the articles where we could determine the type of distribution, it appeared that a wide range of different priors were used (e.g., beta-binomial, Cauchy, Jeffreys, gamma, Weibull, etc.).

The discussion about the level of informativeness of the prior varied article-by-article and was only reported in 56.4% of the articles. It appears that definitions categorizing “informative,” “mildly/weakly informative,” and “noninformative” priors is not a settled issue. This became clear when we extracted the information about how authors refer to their own priors in their articles; see the word cloud in Figure 4. Note that some authors did mention the priors and/or hyperparameters but refrained from referring to these priors with a label stating the level of informativeness ($n = 11$).

Some level of informative priors was used in 26.7% of the empirical articles. For these articles we feel it is important to report on the source of where the prior information came from. Therefore, it is striking that 34.1% of these articles did not report any information about the source of the prior. The articles that did report the source, reported sources such as previous research or data (19.2%), pragmatic or logical reasons (8.4%), expert opinions ($n = 2$), or the data itself (i.e., Empirical Bayes methods; $n = 2$).

An important technical aspect of specifying priors is whether the priors are conjugate. Conjugacy is the mathematical property whereby the combination of the prior distribution with the observed data likelihood yields a posterior distribution of the same family of distribution as the prior distribution. This ensures that the posterior has a closed-form expression, which simplifies estimation procedures. Often, an improper prior will be used as a non-informative prior for variance terms. An improper prior is a probability distribution that does not sum or integrate to one. Because it does not integrate or sum to one, it can technically not serve as a probability distribution. Only five of the articles discussed conjugacy in detail (e.g., Ozechowski, 2014). In 22.8% of the articles, all of the equations were provided. In another 22.2% the distributional form of the priors were provided so that the conjugacy can be checked. Only 12 articles used uniform priors and most of the other articles used (multivariate) normal priors for most of the parameters of interest.

Overall, there is a wide variation of priors used in empirical articles. However, it is striking that so many articles were not completely transparent about the priors, the hyperparameters and the level of informativeness of the prior used, as well as the source for the informative priors.

Sensitivity analysis. If the prior is informative, and especially when combined with small data, it might be useful to report a sensitivity analysis. A sensitivity analysis can be helpful to illustrate how robust final model results are when priors are slightly (or even greatly) modified; this provides a better understanding of the

⁷ Based on information retrieved from <http://www.mrc-bsu.cam.ac.uk/software/bugs/>



Figure 4. Wordcloud showing terms used to describe the level of informativeness of the priors in the empirical regression-based articles.

role of the prior in the analysis. For more information about the impact of priors on the results see Depaoli (2012, 2013) or van de Schoot et al. (2015).

Based on the wording used by the original authors of the articles, as reported above 30 empirical regression-based articles used an informative prior. Of those, 12 (40%) reported a sensitivity analysis; only three of these articles fully described the sensitivity analysis in their articles (see, e.g., Gajewski et al., 2012; Matzke et al., 2015). Out of the 64 articles that used uninformative priors, 12 (18.8%) articles reported a sensitivity analysis. Of the 73 articles that did not specify the informativeness of their priors, three (4.1%) articles reported that they performed a sensitivity analysis, although none fully described it.

In contrast, there are also times when a sensitivity analysis is not necessary when informative priors are implemented. Take the case of the approximate zero for cross loadings or testing for measurement invariance CFAs, as introduced in Muthén and Asparouhov (2012; see also Moore, Reise, Depaoli, & Haviland, 2015; van de Schoot et al., 2013), the informed prior is a specific setting in order to create the approximate zero's. In this case, the informed prior is not really a "subjective" prior and a sensitivity analysis is not necessary because this prior is being used in a confirmatory sense as a model restriction. However, as showed by several authors who did perform a sensitivity analysis in such situations (e.g., Chiorri, Day, & Malmberg, 2014; Ciecuch, Davidov, Schmidt, Algesheimer, & Schwartz, 2014; van de Schoot et al., 2013) the posterior estimates of the latent means are influenced by different specifications for the prior variances imposed on the approximate zero priors.

In all, we hope that this information helps researchers to understand the importance of including a sensitivity analysis in relevant applied contexts. Understanding the impact of the prior and robustness of results is key in many empirical situations

implementing Bayes, and sensitivity analysis is a tool that can help to derive this information.

Convergence. Examining and reporting on chain convergence is very important in Bayesian statistics because results are only trustworthy if the postburn-in portion of the MCMC chain(s) truly converged. In 43.1% of the empirical articles, we could not find any information on how (or whether) convergence was assessed. In 23.4% of the articles, convergence was only implicitly reported but not directly assessed. For example, in an article it may be stated that thousands of iterations were reported, or that only a visual check was performed. In cases where convergence statistics were reported, the Gelman-Rubin criterion (Gelman & Rubin, 1992) was most often reported (26.9%). In fewer articles (6.6%), other convergence criteria like the Geweke criterion (Geweke, 1991), were reported.

Overall, we recommend that convergence criteria always be reported to ensure that results are viable for each model parameter estimated.

Model fit. To quantify model fit in the context of posterior predictive checking, posterior predictive p values (ppp-values) can be computed (Gelman, Carlin, Stern, & Rubin, 2004). The model test statistic, the chi-square value, based on the data is then compared to the same test statistic computed for simulated (future) data. Then, the ppp-value is defined as the proportion of chi-square values obtained in the simulated data that exceed that of the actual data. Ppp-values around .50 indicate a well-fitting model. Ppp-values were reported in 32 articles (19.2%), but it highly depends on the statistical technique used whether the ppp-value is (by default) available. For example, 40.9% of the SEM articles reported the ppp-values, whereas 96% of the articles using analysis of variance did not report the ppp-value.

Overall, the posterior predictive p value is not used as a standard tool to evaluate the fit of a model.

Model selection. Within the Bayesian framework typically two ways of model selection is used: (a) by means of the Bayes Factor (BF; Kass & Raftery, 1995), which was used in 18.5% of the articles (but in 60.9% of the analyses of variance articles); and (b) the deviance information criterion (Spiegelhalter, Best, Carlin, & van der Linde, 2002) which is reported in 22.8% of the articles (but in 38.6% of the SEM articles). We discuss both in some more detail below.

Bayes factors. The first way in which BFs can be applied, is to use them as an alternative to classical null hypothesis testing (e.g., Morey & Rouder, 2011; Rouder, Morey, Speckman, & Province, 2012; Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010; Wetzels et al., 2011), which also includes applications explicitly assessing the level of support of the null hypothesis (Annis, Lenes, Westfall, Criss, & Malmberg, 2015; Matzke et al., 2015). The second type of BFs are used for evaluating informative hypotheses, with many technical articles (e.g., Hoijtink, 2001; Klugkist, Laudy, & Hoijtink, 2005; Mulder et al., 2009), tutorials (e.g., Hoijtink, Béland, & Vermeulen, 2014; Klugkist, Van Wesel, & Bullens, 2011; van de Schoot et al., 2011), and applications (e.g., Van Well, Kolk, & Klugkist, 2008; Wong & van de Schoot, 2012) published.

Many researchers argue that BFs are to be preferred over p values, but as stated by Konijn, van de Schoot, Winter, and Ferguson (2015) potential pitfalls of a Bayesian approach include BF-hacking (cf., "Surely, God loves a Bayes Factor of 3.01 nearly

as much as a BF of 2.99”). This can especially occur when BF values are small. Instead of relying on a single study to draw substantive inferences, we advocate that replication studies and Bayesian updating are still necessary to draw conclusions (see also, Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012). A good example of “updating” is the article of Van Den Hout et al. (2014), who reanalyzed data from Gangemi, Mancini, and van den Hout (2012) to compute Bayes Factors, after which they replicated the original study and updated the original Bayes Factors with new findings. There are more authors who used the idea of updating (see also Boekel et al., 2015; Donkin, Averell, Brown, & Heathcote, 2009; Hasshim & Parris, 2015; Montague, Krawec, Enders, & Dietz, 2014; Neyens, Boyle, & Schultheis, 2015), for example, they aimed to replicate results across multiple data sets within one study (de Leeuw & Klugkist, 2012; Dry, 2008), or they used Study 1 as prior information for Study 2 within one article (Milfont & Sibley, 2014). Furthermore, Annis et al. (2015) reanalyzed a study by Dennis, Lee, and Kinnell (2008), itself a Bayesian application, and showed that their findings were actually inconclusive and debatable. Van de Schoot et al. (2014) concluded, after updating the results over four data sets, that updating prior knowledge with new data leads to more certainty about the outcomes of the final analyses and brings more confidence in the conclusions.

Bayes factors as an alternative to hypothesis testing is advocated in many articles. However, this approach is not yet prevalent in empirical articles. It appears from our search that there are currently more tutorials on this topic compared to actual applications. However, we believe that this trend will shift and that the field will begin to see more applications implementing BFs in this manner, including “updating” articles.

Deviance information criterion. The DIC can be used to compare competing models, similar to the AIC (Akaike, 1981) and BIC (Schwarz, 1978). The posterior DIC is proposed in Spiegelhalter et al. (2002) as a Bayesian criterion for minimizing the posterior predictive loss. It can be seen as the error that is expected when a statistical model based on the observed dataset is applied to a future dataset. The loss function of a future dataset is given the expected posteriori estimates of the model parameters based on the observed dataset. If it was possible to know the true parameter values, then the loss function could be computed. However, because these are unknown, the DIC takes the posterior expectation where the first term is comparable with the fit part of the AIC and BIC. The second term is often interpreted as the “effective number of parameters,” but is formally interpreted as the posterior mean of the deviance minus the deviance of the posterior means. Just like with the AIC and BIC, models with a lower DIC value should be preferred and indicates the model that would best predict a replicate dataset, which has the same structure as that currently observed.

Tutorial and Software Articles

One important component we discovered during this systematic review was that many different types of tutorial articles have been published on a variety of Bayesian features and software programs. In order to provide a concise list of these resources, we have created a table in the online material breaking down the 100 tutorial articles by content area and listing the relevant references.

We have also done this for Bayesian-specific software programs that have been described in the literature. We acknowledge that our search was not specifically geared toward software, so it is probable we are missing some developments in this area that occurred outside of (or even within) psychology. However, we still hope this index can be useful for new users of Bayesian methods.

Trends Across Categories

Bayes’s Theorem in Cognitive Psychology

Bayes’s theorem plays a major role in the field of cognitive psychology in various ways. See Lee’s (2011) article published in a special issue of the *Journal of Mathematical Psychology* for an excellent overview of the ways in which Bayes’s theorem is used within cognitive psychology. In short, there are three main ways Bayes’s theorem is currently applied within cognitive psychology. First, it is used as a theoretical framework for how the mind makes inferences about the world. These models are strictly used as a theoretical explanation for human reasoning and behavior. It may seem counterintuitive, but the data are still analyzed with traditional, frequentist methods. Lee states that the Bayesian framework has gained in popularity since the start of the 21st century. Our findings support this conclusion (see Figure 5A, green area). While there were some articles published before the year 2000, this application of Bayes’s theorem really took off from 2003 onward. Some recent examples of this application of Bayes’s theorem are: Gopnik, Griffiths, and Lucas (2015); Juslin, Nilsson, Winman, and Lindskog (2011); and Kinoshita and Norris (2012; see also Table 1, “Bayes as cognition/learning”).

A second use of Bayesian methods in cognitive psychology is through Bayesian hierarchical modeling. These models attempt to relate models of psychological processes to actual data. Bayesian hierarchical modeling allows the researcher to create a detailed, concrete model of the processes that are assumed to be part of the human thinking process and to compare predictions of behavior from this model with observed data of actual human behavior. If the model is able to adequately predict the actual data, then this tells us something about the mechanisms that precede the observed behavior. As Figure 5A shows (blue area), the use of these computational hierarchical Bayesian models has increased after 2006, closely following the trend of Bayes’ as a theoretical framework for cognition. Some recent examples of this application of Bayes’s theorem are: Ferreira, Castelo-Branco, and Dias (2012); Lee and Sarnecka (2011); and Scheibehenne and Studer (2014; see also, Table 1, “Computational model”).

A third use of Bayesian methods in cognitive psychology is through the direct application of Bayesian statistics on observed data. Lee (2011, p. 1) concludes: “It seems certain that Bayesian statistics will play a progressively more central role in the way cognitive psychology analyzes its data.” In our search, we have indeed found some articles that directly apply Bayesian statistics to cognitive data. For example, Andrews, Vigliocco, and Vinson (2009) used a Bayesian ANOVA to analyze various computational models based on cognitive theory about how humans learn semantic representations. Voorspoels, Navarro, Perfors, Ransom, and Storms (2015) applied Bayesian *t* tests to compare two experimental conditions on word-generalization, while Voorspoels, Storms, and Vanpaemel (2012) used a hier-

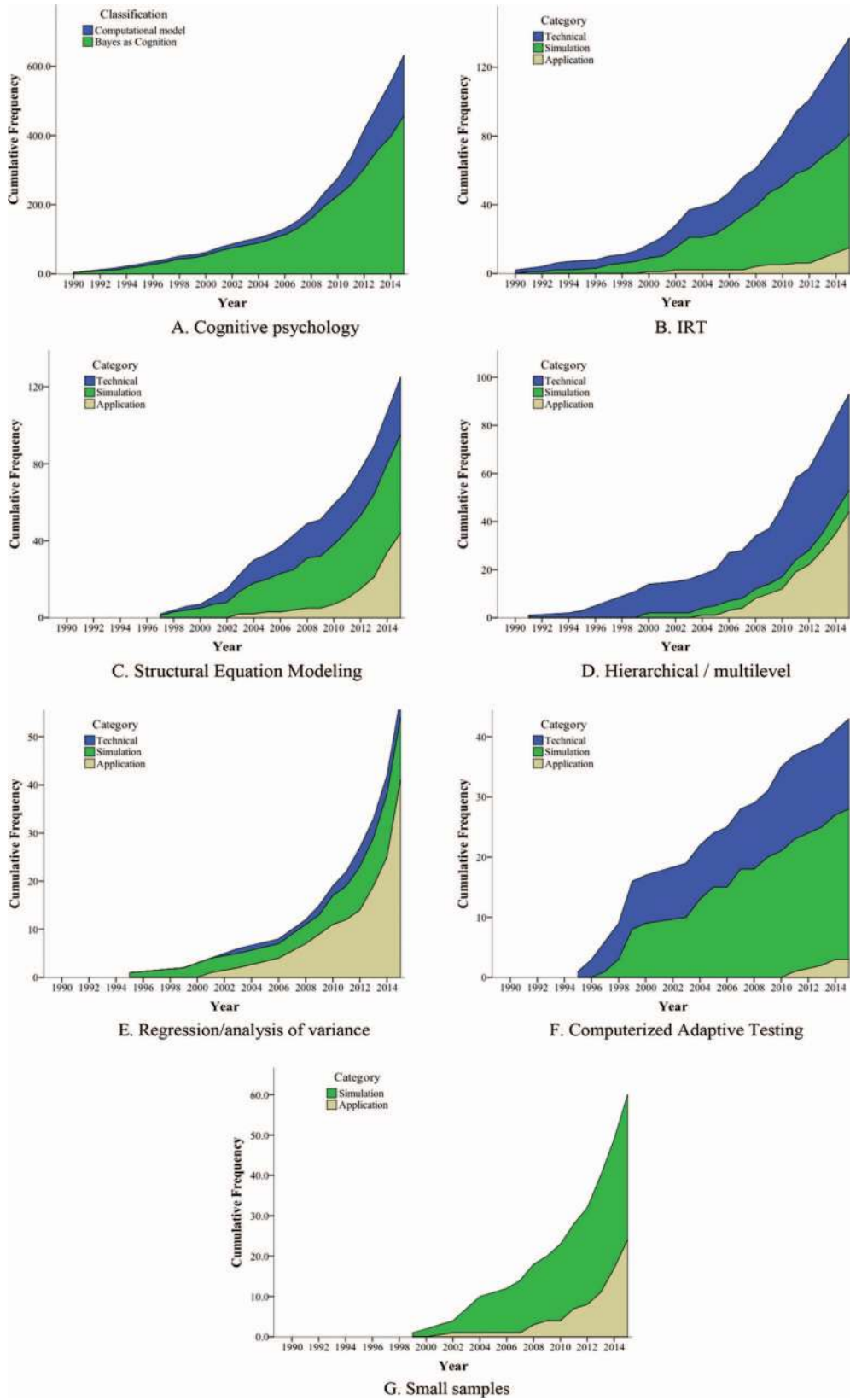


Figure 5. Development of articles using Bayes' using various statistical models/techniques. See the online article for the color version of this figure.

archical Bayesian model and Bayes Factors to investigate contrast effects in category learning. Furthermore, [Steyvers, Lee, and Wagenmakers \(2009\)](#) used Bayes Factors to compare four models of learning, of which one was based on the theoretical “Bayesian learner.” Finally, [Sims, Neth, Jacobs, and Gray \(2013\)](#) created a rational-learner model and used Bayesian methods to investigate melioration, or choosing a lesser, short-term gain over a greater, long-term gain. These articles are included under “Regression-based” articles in [Table 1](#).

In conclusion, Bayes’s theorem plays an important role in cognitive psychology, and its main application is slowly changing from a theoretical framework to an actual method for estimating models.

Item Response Theory

Within the category “regression-based models,” IRT has arguably been the largest area within psychology (and related fields) where Bayesian statistics has surfaced. The use of MCMC for estimating IRT-based models dates back at least to the technical articles of [Albert \(1992\)](#); [Jannarone, Yu, and Laughlin \(1990\)](#); and [Tsutakawa and Johnson \(1990\)](#). More accessible articles on IRT include [Geerlings, Laros, Tellegen, and Glas \(2014\)](#); [Ip \(2010\)](#); [Sinharay \(2005\)](#); [Wainer, Wang, Skorupski, and Bradlow \(2005\)](#); and [Wang, Chen, and Jin \(2015\)](#). Out of all of the technical articles we discovered, 30.3% were based on IRT modeling, making it the largest modeling group represented in this category of technical articles. IRT also had a large presence in the Bayesian simulation literature. One-third of all simulation articles were based on IRT models. Only 8.9% of the applied Bayesian articles used IRT techniques. These trends are displayed in [Figure 5B](#), where it can be seen that a larger number of Bayesian IRT articles were published in all three categories (theoretical, simulation, and application) as time progressed.

It is surprising that the relatively large proportion of Bayesian technical and simulation articles dealing with IRT was not mimicked in the applications published.

Structural Equation Modeling

[Figure 5C](#) presents trends for theoretical, simulation, and application articles implementing Bayesian SEM. Within the technical articles (16.2%) and the simulation articles (25.8%), SEM is the second most common type of statistical model used following closely behind IRT models. However, within applications it is the number one statistical model used (26.0%). When inspecting [Figure 5C](#), we can see that there is a steep increase in the number of articles published in all three of these categories over time. However, it is striking that the number of Bayesian SEM applications has experienced a dramatic spike since about 2012. This is in contrast with the technical and simulation articles, which have been published at a relatively steady rate over the last several years of the assessment.

It appears that the presence of Bayesian SEM continues to grow in all three areas, but that we may expect to see a faster increase in the number of applications in the coming years.

Hierarchical/Multilevel

The next most common statistical method within technical (13.5%) and empirical articles (26%) is multilevel modeling (excluding hierarchical IRT models; these are categorized in IRT; see [Figure 5D](#)). We found hardly any simulation studies (3%) on this topic; we suspect that the simulation studies on Bayesian multilevel modeling are published outside the field of psychology, see for example [Hox, van de Schoot, and Matthijsse \(2012\)](#). Note that in the literature, multilevel modeling is also called hierarchical modeling, random effects modeling, variance component modeling, contextual analysis, mixed linear modeling, and many other variations. All of these forms make use of the fact that the data have a nested structure.

In all, there is a steady increase of articles using Bayesian multilevel modeling in an application context.

Regression/Analysis of Variance

A notable portion of empirical articles used univariate linear regression models (10.7%) or analysis of variance (13.6%), while hardly any of the technical (1.1%) or simulation articles (5.6%) we found focused on these statistical models (see [Figure 5E](#)). One possibility for this lack of technical and simulation articles could be the timing of our review. Regression and analysis of variance models are decidedly less-complex and more established in the literature compared with some of the other models we have reported on. It could be that the development of Bayesian estimation for these models occurred in the literature predating our search years.

Computerized Adaptive Testing

A field that received a growing amount of attention in technical (8.1%) and simulation articles (12.6%) is CAT (see [Figure 5F](#)). Bayesian applications of CAT, however, are sparsely published (1.8%). We found only three such articles in our search: [Jones \(2014\)](#) published a full Bayesian application, whereas [O’Donnell and Colombo \(2011\)](#) and [Tong, Mao, and Goldreich \(2013\)](#) merely used CAT in the methods section.

In conclusion, Bayesian methods have been (sparsely) applied to CAT for many years, and we expect that trend to continue.

Small Samples

The focus of using Bayesian methods in small sample contexts is increasing for simulation articles, as well as the number of regression based applications (see [Figure 5G](#)). There is a certain appeal of Bayesian methods to researchers who only have access to a relatively small number of participants or cases. In general, the more information a researcher can specify (via the prior) before seeing the data, the smaller the sample size needed to obtain the same certainty compared to an analysis without specifying any prior knowledge. Or as stated by [Rouder, Sun, Speckman, Lu, and Zhou \(2003, p. 589\)](#):

... for small sample sizes, there is an occasional tendency for the ML estimates to be unreasonably extreme. In contrast, by borrowing strength across participants, Bayes estimation “shrinks” extreme estimates. The results are that the Bayes estimators are more accurate than the corresponding ML estimators.

An attractive feature is that priors can be used in a variety of ways: to have an influence, a tiny influence, or virtually no influence at all; as is nicely summarized by Baldwin and Fellingham (2013, p. 162):

Even if researchers are comfortable with a prior distribution, they may insist on using flat or diffuse priors. As the simulation study demonstrated, when sample sizes are small flat priors can influence results and lead to additional bias and decreased efficiency. Indeed, diffuse priors can be unnecessarily inefficient if the priors imply that mathematically possible but extreme outcomes are equally probable to nonextreme values or if the priors ignore existing evidence. Consequently, we recommend that researchers devote the necessary time and effort to construct plausible, thoughtful priors that incorporate existing information about a research area. Like any aspect of the research process, the details of the prior distributions as well as the methods used to select them should be well described and justified in a research report.

We found 37 simulation articles that examined the performance of Bayesian methods in small samples; this represented 18.7% of the Bayesian simulation articles identified in the search. The types of models these articles focused on were the following: SEM (45.9%), IRT (18.9%), multilevel modeling and CAT were tied (8.1%), regression (2.7%), and other miscellaneous model types (16.2%). Within these articles, the performance of Bayes varied depending on the sample size. However, many of the articles reported improvements relative to, for example, ML estimation when sample sizes were small. Note that “small” sample sizes varied drastically depending on the models being examined. For example, Kiefenbeld and Natesan (2012) found that the graded response IRT model improved in parameter recovery with samples as small as 75. Mediation model analyses reported sample sizes as small as 25 and still found an improvement (Yuan & MacKinnon, 2009). Improved estimates were also reported in multilevel CFAs with only 10–20 clusters (Hox, Moerbeek, Kluytmans, & van de Schoot, 2014; Hox et al., 2012).

It is not surprising that we found many simulation articles stating that informative priors improve parameter estimates to a greater degree than uninformative (or diffuse) priors. This finding is likely because the informative priors usually contained correct information about the parameter. We found nine simulation articles detailing how informative priors aided in proper estimation (e.g., accuracy and efficiency) for a variety of model-types. Within the simulation studies, it is emphasized that prior distributions should be composed with care (as is argued in, e.g., Gao & Chen, 2005), because small sample results are so easily affected. In this respect, Baldwin and Fellingham (2013) also suggested that researchers invest time and effort to incorporate existing information into prior distributions. They found that flat and diffuse priors can affect results, increase bias, and reduce efficiency. As is shown in Lee, Song, and Tang (2007), among several other articles, specifying inaccurate priors for small data can result in estimates not as good as those that were obtained with the more accurate prior inputs. Thus, priors should be chosen with care, but luckily already even mildly informative priors can be used effectively to obtain estimates for small samples. As argued by Natesan (2015)

There is no magical cure for lack of information. However, when a researcher is faced with small sample ordinal data Bayesian estimation provides reasonable point and interval estimates.

We found 18 empirical articles that cited using Bayesian methods for small sample size issues. The range of sample sizes we found were: 23 for an SEM model (Ozechowski, 2014), and up to 1,000 for a mixed logit model (Choi et al., 2008). The majority (42.1%) of them used uninformative or weak priors, with another 26.3% using software default settings which are also uninformative in most (if not all) cases. Some articles (10.5%) did not mention the informativeness of the priors, and empirical Bayes, informative priors, hyperpriors, and informative but weak priors were each used in 5.3% of the articles.

In sum, given the strong impact of priors in small sample situations, as demonstrated by the simulation studies, small sample studies should be especially clear about specific hyperparameters. In addition, sensitivity analyses are highly recommended.

To examine whether there are follow-up efforts for replication or cross-validation of articles that cite small sample size as the reason to adopt a Bayesian approach, we used the following method. First, we searched for the original article through Google Scholar, and then we selected the articles that cited the original article ($n = 246$ citations). Within this list, we then searched for articles mentioning “replic*” somewhere in the text. We subsequently read all of the flagged articles ($n = 68$ citations, linked to 12 of our small sample articles identified) to see whether they were attempting to replicate the original article.

Our findings show that only three of the small sample articles were used as the basis for a replication study. In all cases, it concerned a conceptual replication. First, the study by Freisthler and Weiss (2008) was conceptually replicated in an unpublished manuscript (Freisthler & Martin, 2010). The authors write in their discussion: “This is the first study to attempt to replicate findings of outlet density and maltreatment in a state other than California and the first to do so in a largely rural state.” (p. 11). Second, the study by Kunreuther, Silvasi, Bradlow, and Small (2009) was conceptually replicated in the same year by Gong, Baron, and Kunreuther (2009). The replication study found substantively different results regarding the outcome of cooperation in a gaming situation. Third, the study by Imel et al. (2011) was conceptually replicated by Owen et al. (2015). The authors write in their discussion: “[t]he current study replicated and extended findings. . . . These findings are consistent with the previous studies examining counselor variability in racial/ethnic disparities” (p. 7).

Given the recent call on increasing replication within psychology (Asendorpf et al., 2013), we reiterate that this is also an important topic to address within Bayesian statistics especially when sample size are small.

Controversy Between Bayesians and Frequentists

Finally, our search uncovered several different articles written from a Bayesian viewpoint that seek to criticize frequentist methods. Although Bayesians and frequentist statisticians have been publishing in parallel to one another for decades, there has been a good deal of crossover-criticism from both sides. These branches of statistics have a history rooted in controversy due to fundamental differences in philosophical views. This controversy is nicely illustrated by the following two quotes:

As embarrassing as it is to us as Bayesians, we have effectively derived an orthodox null hypothesis test for feature coherence. We apologize. (Navarro & Perfors, 2010, Footnote 7)

Besides correcting the most obvious flaws of NHST in a manner reminiscent of how meta-analysis does it, Bayesian statistics can be seen as providing answers to the questions researchers would be asking, unless they had first been taught a flawed method. (Sohlberg & Andersson, 2005, p. 69)

Occasionally, the tension between these sides of statistics results in a series of publications. One example resides in the discussion between the research group led by Bem (Bem, 2011; Bem, Utts, & Johnson, 2011) and the group of Wagenmakers, Wetzels, Borsboom, and van der Maas (2011). Bem (2011) originally published a study on nine experiments, where he attempted to demonstrate the existence of ψ (future events retroactively affecting people's responses) using frequentist statistics. Wagenmakers et al. (2011) reanalyzed their data using a Bayesian t test and concluded that the evidence in favor of ψ was weak-to-nonexistent. Bem et al. (2011) reanalyzed the data again by claiming Wagenmakers et al. (2011) used an unrealistic prior distribution for their analyses; Wagenmakers et al. (2012) replied again and also other Bayesians joined the discussion (dis)agreeing with the methods used not only in the original Bem study, but also used in the Wagenmaker article—see Hoijtink, van Kooten, and Hulsker (2016b); Morey, Wagenmakers, and Rouder (in press); Hoijtink, van Kooten, and Hulsker (2016a), and Stern (2016)—and so the story continues.

These examples highlight the continued debate over philosophical issues underlying frequentist and Bayesian methods and within the Bayesian field, such as the impact and use of prior knowledge within the estimation process.

Discussion

This is an exciting time, where we can watch the field of applied statistics change more than ever before. The way in which researchers think about and answer substantive inquiries is slowly taking on a new philosophical meaning that now incorporates previous knowledge and opinions into the estimation process. We began this article with a quote from Kruschke (2011; as forecasted by Lindley) indicating that we would likely see a rise in Bayesian applications in this century. Our systematic review attempted to assess whether this statement was valid, and if a rise in the use of Bayesian methods was evident based on publication patterns. Our findings showed that there was some merit in this thought. In fact, the use of Bayesian methods in applied Psychological work has steadily increased since the nineties and is currently taking flight. It was clear in this review that Bayesian statistics is used in a variety of contexts across subfields of psychology and related disciplines. There are many different reasons why one might choose to use Bayes. For example, one might use Bayesian methods for: the use of priors, estimating otherwise intractable models, circumventing small sample size issue, and so forth. Some researchers may even choose to use Bayesian methods for modeling uncertainty in the model (e.g., through Bayesian model averaging). Ultimately, the use of Bayes has increased and broadened in the sense that this methodology can be used in a flexible manner to tackle many different forms of questions.

The relatively slow growth in popularity of Bayesian methods up to around the year 2010 might be due to several reasons. First, the traditional statistical approach being taught to psychologists is to think in terms of the null hypothesis. Psychologists have largely been married to their null hypotheses, and they often design studies in terms of a point-null hypothesis. Moreover, many psychologists make use of experiments in combination with rather straightforward analyses, for which you may not need to leave the frequentist environment. Fields that implement very complex optimization/integration problems are more prone to the Bayesian concept as another way to approach statistical problems, making Bayesian methods more of the norm in those fields. All of this is to say that there may be many different reasons why psychologists have seemingly lagged behind in the use of Bayesian statistics compared with some fields. It could be that psychologists do not readily identify situations in which the use of Bayesian methods may be beneficial. Relatedly, it might be that it is rather *involved* to actually use Bayesian methods for the uninitiated researcher, or that Bayesian statistics as a practice is not user-friendly.

Using Bayesian methods also requires the generation of knowledge, and many psychological researchers may not be familiar thinking this way. For example, thinking of parameters in terms of distributions, or specifying priors, are topics many users may not usually consider during statistical analysis. We feel there has traditionally been a lack of exposure to Bayesian statistics within applied psychological fields. This issue of exposure can, in part, be solved to some degree with building user-friendly software. However, even then, as you move away from simple models with uninformative priors, the user has to have more in-depth knowledge of the Bayesian process. In other words, it is a combination of making Bayes accessible, as well as educating future generations of researchers about these methods in detail.

In this sense, we are faced with a “catch-22” as we move forward. Bayes gets more interesting as you move toward more complex modeling contexts, and these contexts are advantageous to use as teaching tools to show the power and impact of Bayesian statistics. However, these sorts of situations are also linked to more complicated software components, which make it much harder to start with for the novice Bayesian researcher. Luckily there is a steady increase of very accessible textbooks (see, e.g., Gill, 2014; Kaplan, 2014; Kruschke, 2014; Lunn, Jackson, Best, Thomas, & Spiegelhalter, 2012; Lynch, 2007), articles gently introducing Bayes (see the online material for 107 of such articles), checklists on what to report (Depaoli & van de Schoot, 2015; Pullenayegum, Guo, & Hopkins, 2012; Spiegelhalter, Myles, Jones, & Abrams, 2000; Sung et al., 2005), as well as courses offered all around the world that assist researchers in becoming *involved*.

These new resources aimed at being user-friendly for novice users might explain the huge increase of empirical articles published since 2010 (and especially applications since 2014). That is, training in Bayesian methods appears to have slowly become a standard part of the statistical toolkit of researchers working in the field of psychology. From an educational perspective, universities should develop a statistics curriculum that more evenly includes frequentist *and* Bayesian methods, giving students the option of working with both early on and deciding for themselves which method they would like to apply in their own research.

We feel education and exposure is the single biggest issue related to Bayesian statistics in psychology. Although the use of

Bayesian methods is on the rise, we (the field) also need to continue to develop our understanding of how methods perform. Our findings (see Figure 4) show that technical articles, simulation articles, and applications are not published at the same rate. For several subdisciplines, technical and simulation articles are being published at a faster rate compared to applications. However, this pattern of results is typical for methods being newly adopted into a field. This pattern is typically found in fields where Bayes is seen as a cognitive process, which is translated into a Bayesian computational model. However, in these cases, most of the actual data analyses are done with frequentist statistics. We also see a higher rate of Bayesian technical and simulation articles being published in the fields of IRT and CAT. This is interesting because it shows that the methodological literature is out-pacing the applied literature. It may be that there is relatively more methodological interest in Bayes in psychology compared with applications. We encourage experts in Bayesian statistics to support/coauthor applications. When they do so, they should become role models in adequate reporting. Producing more quality applications within different subfields of psychology will help to broaden the usage and demonstrate proper implementation of Bayesian methods.

Bayesian statistics follows its own rules since there are elements included in the Bayesian framework that are fundamentally different from frequentist settings. Perhaps one of the more alarming findings from our review was that there appears to be a low proportion of articles published that properly report important issues surrounding Bayesian estimation (e.g., the source and specification of the priors, etc.). In any science, it is our duty to ensure that results are correct, reported accurately, and (perhaps most importantly) reproducible. The absence of information regarding convergence and the specification of priors in studies utilizing Bayesian MCMC-based methods may greatly hinder these goals. It is possible that as the use of Bayes increases, the reporting of certain details may decrease. For example, in articles that use ML estimation, proper convergence of the estimation algorithm is usually assumed and not reported. However, the nature of current Bayesian MCMC-based methodology is such that reporting convergence (visually and through diagnostics) is still arguably quite important, especially because it is much more difficult a task of assessing convergence in Bayes than converging to a point estimate (as in frequentist methods).

Furthermore, reporting on priors will always be important because priors can fundamentally change the substantive results obtained through the posterior. Unfortunately, there appears to be a trend in psychology where some of the key features of Bayesian methodology are either not being properly screened for, or perhaps intentionally omitted at the behest of journal. This can only lead to a decrease in the reproducibility of reported results, and in an increase of confusion about the results that are reported using Bayesian methods. For researchers concerned with reporting all relevant details for their Bayesian analysis, see the online material for a list of 50 articles we feel presented adequate details regarding Bayesian estimation and results (see Depaoli & van de Schoot, 2015 for a related checklist for the reporting of Bayesian analysis). While many editors and authors may omit such details in the interest of manuscript length, there is no excuse in modern times for not making such details available online—ideally as part of the

journal website or an official data storage with a DOI number. In the online supplementary materials we provide a list of 45 articles, which can serve as inspiration for how Bayesian statistics could be reported. Most articles reported all specific information about priors, convergence, and sensitivity analyses in the main text. Others published an online appendix with this detailed information.

Although many argue that Bayesian statistics solve issues encountered in frequentist statistics, Bayesian analyses are not immune to phenomena related to the replication crisis in science, which include questionable research practices (QRPs), or *p*-hacking behavior. For example, to evaluate whether a BF can be considered substantial, cut-off values for minimum BF-values are proposed by Kass and Raftery (1995), but these values are not meant as strict cut-off values. This issue is not unlike how $p < .05$ is often (mis)used. We argue strongly against using unjustified cut-off values as decision rules within the Bayesian framework because this might result in similar hacking-behavior as with *p* values, called BF-hacking by Konijn et al. (2015). They state that a BF of 3.01 should not be considered substantial compared with a BF of 2.99, which would then be “not worth more than a bare mention” (cf. Kass & Raftery, 1995). Konijn et al. (2015) “adjusted” the famous quote of Rosnow and Rosenthal: “. . . surely, God loves the .06 nearly as much as the .05” (Rosnow & Rosenthal, 1989, p. 1277), to the Bayesian framework: “God would love a Bayes Factor of 3.01 nearly as much as a BF of 2.99.”

Study Limitations

Of course no study is without limitations. It is imaginable that we have missed Bayesian articles in our review, simply because they do not mention our search terms in their title, abstract, or keywords. Another limitation could also be that we only used Scopus as a search engine and could have therefore missed out on some viable articles. Future research could possibly adjust the search method applied by Rietbergen (under review), who used a piece of software capable of searching within the article text via the PubMed search engine. On the other hand, with our search terms, we were as inclusive as possible and were able to identify thousands of relevant articles; we even covered articles that did not actually use any Bayesian methods in their study but simply mentioned “Bayes.”

Another issue is that some of the trends detailed here may not be unique to Bayesian statistics, but rather may reflect overall changes in methodology. For example, we found that open access journals are growing in Bayesian statistics, but they may also be growing in popularity for other general statistical areas (Björk et al., 2010). Given that our focus was strictly on the trends in Bayesian statistics, we are not able to make full comparisons with trends outside of Bayesian statistics. However, the trends that we did uncover are key to understanding the role that Bayesian methods have been playing in the psychological literature.

Concluding Thoughts

In conclusion, we are optimistic about the increased use of Bayesian methods. Much research has shown that there are important benefits to using this estimation approach in psychological inquiries. Bayesian statistics in psychology, just like in

many other fields, is becoming a true alternative to frequentist methods. State of the art Bayesian features are increasingly used to explore new research possibilities, to seek out new ideas, to answer new research questions, and to boldly go where just few researchers have gone before. We would like to challenge psychological researchers to move beyond frequentist methods, which were the dominant research tool of the 20th century, and move toward the new statistics of the 21st century. We believe that if applied researchers tend to the problematic issues found in this review, future Bayesian work can be more thorough, transparent, and replicable.

References

References included in the systematic review can be located in the supplementary materials.

- Ahn, W. Y., Vasilev, G., Lee, S. H., Busemeyer, J. R., Kruschke, J. K., Bechara, A., & Vassileva, J. (2014). Decision-making in stimulant and opiate addicts in protracted abstinence: Evidence from computational modeling with pure users. *Frontiers in Psychology, 5*, 849.
- Akaike, H. (1981). Likelihood of a model and information criteria. *Journal of Econometrics, 16*, 3–14. [http://dx.doi.org/10.1016/0304-4076\(81\)90071-3](http://dx.doi.org/10.1016/0304-4076(81)90071-3)
- Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational Statistics, 17*, 251–269. <http://dx.doi.org/10.3102/10769986017003251>
- Albert, M. K. (2000). The generic viewpoint assumption and Bayesian inference. *Perception, 29*, 601–608. <http://dx.doi.org/10.1068/p3050>
- Allen, J. J. B., & Iacono, W. G. (1997). A comparison of methods for the analysis of event-related potentials in deception detection. *Psychophysiology, 34*, 234–240. <http://dx.doi.org/10.1111/j.1469-8986.1997.tb02137.x>
- Almond, R. G., DiBello, L. V., Moulder, B., & Zapata-Rivera, J. D. (2007). Modeling diagnostic assessments with Bayesian networks. *Journal of Educational Measurement, 44*, 341–359. <http://dx.doi.org/10.1111/j.1745-3984.2007.00043.x>
- Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review, 116*, 463–498. <http://dx.doi.org/10.1037/a0016261>
- Annis, J., Lenes, J. G., Westfall, H. A., Criss, A. H., & Malmberg, K. J. (2015). The list-length effect does not discriminate between models of recognition memory. *Journal of Memory and Language, 85*, 27–41. <http://dx.doi.org/10.1016/j.jml.2015.06.001>
- Arminger, G., & Muthén, B. O. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the metropolis-hastings algorithm. *Psychometrika, 63*, 271–300. <http://dx.doi.org/10.1007/BF02294856>
- Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J., Fiedler, K., . . . Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality, 27*, 108–119. <http://dx.doi.org/10.1002/per.1919>
- Ashby, D. (2006). Bayesian statistics in medicine: A 25 year review. *Statistics in Medicine, 25*, 3589–3631. <http://dx.doi.org/10.1002/sim.2672>
- Baldwin, S. A., & Fellingham, G. W. (2013). Bayesian methods for the analysis of small sample multilevel data with a complex variance structure. *Psychological Methods, 18*, 151–164. <http://dx.doi.org/10.1037/a0030642>
- Bayes, M., & Price, M. (1763). An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London, 53*, 370–418. <http://dx.doi.org/10.1098/rstl.1763.0053>
- Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology, 100*, 407–425. <http://dx.doi.org/10.1037/a0021524>
- Bem, D. J., Utts, J., & Johnson, W. O. (2011). Must psychologists change the way they analyze their data? *Journal of Personality and Social Psychology, 101*, 716–719. <http://dx.doi.org/10.1037/a0024777>
- Björk, B.-C., Welling, P., Laakso, M., Majlender, P., Hedlund, T., & Gudnason, G. (2010). Open access to the scientific journal literature: Situation 2009. *PLoS ONE, 5*, e11273. <http://dx.doi.org/10.1371/journal.pone.0011273>
- Boekel, W., Wagenmakers, E. J., Belay, L., Verhagen, J., Brown, S., & Forstmann, B. U. (2015). A purely confirmatory replication study of structural brain-behavior correlations. *Cortex, 66*, 115–133. <http://dx.doi.org/10.1016/j.cortex.2014.11.019>
- Bujacz, A., Vittersø, J., Huta, V., & Kaczmarek, L. D. (2014). Measuring hedonia and eudaimonia as motives for activities: Cross-national investigation through traditional and Bayesian structural equation modeling. *Frontiers in Psychology, 5*, 984.
- Bulbulia, J. A., Xygalatas, D., Schjoedt, U., Fondevila, S., Sibley, C. G., & Konvalinka, I. (2013). Images from a jointly-arousing collective ritual reveal affective polarization. *Frontiers in Psychology, 4*, 960.
- Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., . . . Riddell, A. (in press). Stan: A probabilistic programming language. *Journal of Statistical Software*.
- Cavanagh, J. F., Wiecki, T. V., Kochar, A., & Frank, M. J. (2014). Eye tracking and pupillometry are indicators of dissociable latent decision processes. *Journal of Experimental Psychology: General, 143*, 1476–1488. <http://dx.doi.org/10.1037/a0035813>
- Chiorri, C., Day, T., & Malmberg, L. E. (2014). An approximate measurement invariance approach to within-couple relationship quality. *Frontiers in Psychology, 5*, 983.
- Choi, J. Y., Koh, D., & Lee, J. (2008). Ex-ante simulation of mobile TV market based on consumers' preference data. *Technological Forecasting and Social Change, 75*, 1043–1053. <http://dx.doi.org/10.1016/j.techfore.2007.10.001>
- Cieciuch, J., Davidov, E., Schmidt, P., Algesheimer, R., & Schwartz, S. H. (2014). Comparing results of an exact vs. an approximate (Bayesian) measurement invariance test: A cross-country illustration with a scale to measure 19 human values. *Frontiers in Psychology, 5*, 982.
- Cipora, K., & Nuerk, H. C. (2013). Is the SNARC effect related to the level of mathematics? No systematic relationship observed despite more power, more repetitions, and more direct assessment of arithmetic skill. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 66*, 1974–1991. <http://dx.doi.org/10.1080/17470218.2013.772215>
- de Leeuw, C., & Klugkist, I. (2012). Augmenting data with published results in Bayesian linear regression. *Multivariate Behavioral Research, 47*, 369–391. <http://dx.doi.org/10.1080/00273171.2012.673957>
- Dennis, S., Lee, M. D., & Kinnell, A. (2008). Bayesian analysis of recognition memory: The case of the list-length effect. *Journal of Memory and Language, 59*, 361–376. <http://dx.doi.org/10.1016/j.jml.2008.06.007>
- Depaoli, S. (2012). Measurement and structural model class separation in mixture CFA: ML/EM versus MCMC. *Structural Equation Modeling, 19*, 178–203. <http://dx.doi.org/10.1080/10705511.2012.659614>
- Depaoli, S. (2013). Mixture class recovery in GMM under varying degrees of class separation: Frequentist versus Bayesian estimation. *Psychological Methods, 18*, 186–219. <http://dx.doi.org/10.1037/a0031609>
- Depaoli, S., & van de Schoot, R. (2015). Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist. *Psychological Methods*. Advance online publication. <http://dx.doi.org/10.1037/met0000065>

- Desmet, P., & Feinberg, F. M. (2003). Ask and ye shall receive: The effect of the appeals scale on consumers' donation behavior. *Journal of Economic Psychology, 24*, 349–376. [http://dx.doi.org/10.1016/S0167-4870\(02\)00166-6](http://dx.doi.org/10.1016/S0167-4870(02)00166-6)
- Donkin, C., Averell, L., Brown, S., & Heathcote, A. (2009). Getting more from accuracy and response time data: Methods for fitting the linear ballistic accumulator. *Behavior Research Methods, 41*, 1095–1110. <http://dx.doi.org/10.3758/BRM.41.4.1095>
- Dry, M. J. (2008). Using relational structure to detect symmetry: A Voronoi tessellation based model of symmetry perception. *Acta Psychologica, 128*, 75–90. <http://dx.doi.org/10.1016/j.actpsy.2007.10.001>
- Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70*, 193–242. <http://dx.doi.org/10.1037/h0044139>
- Eizirik, L. M. R., Barbosa, V. C., & Mendes, S. B. T. (1993). A Bayesian-network approach to lexical disambiguation. *Cognitive Science, 17*, 257–283. http://dx.doi.org/10.1207/s15516709cog1702_3
- Fenton, N., Neil, M., & Lagnado, D. A. (2013). A general structure for legal arguments about evidence using Bayesian networks. *Cognitive Science, 37*, 61–102. <http://dx.doi.org/10.1111/cogs.12004>
- Ferreira, J. F., Castelo-Branco, M., & Dias, J. (2012). A hierarchical Bayesian framework for multimodal active perception. *Adaptive Behavior, 20*, 172–190. <http://dx.doi.org/10.1177/1059712311434662>
- Foreman, D., Morton, S., & Ford, T. (2009). Exploring the clinical utility of the Development and Well-Being Assessment (DAWBA) in the detection of hyperkinetic disorders and associated diagnoses in clinical practice. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 50*, 460–470. <http://dx.doi.org/10.1111/j.1469-7610.2008.02017.x>
- Freisthler, B., & Martin, S. E. (2010). *Examining the relationship of alcohol outlet density and child maltreatment in rural areas: The case of Nebraska*. Unpublished manuscript.
- Freisthler, B., & Weiss, R. E. (2008). Using Bayesian space-time models to understand the substance use environment and risk for being referred to child protective services. *Substance Use & Misuse, 43*, 239–251. <http://dx.doi.org/10.1080/10826080701690649>
- Gajewski, B. J., Coffland, V., Boyle, D. K., Bott, M., Price, L. R., Leopold, J., & Dunton, N. (2012). Assessing content validity through correlation and relevance tools a Bayesian randomized equivalence experiment. *Methodology, 8*, 81–96. <http://dx.doi.org/10.1027/1614-2241/a000040>
- Gangemi, A., Mancini, F., & van den Hout, M. (2012). Behavior as information: "If I avoid, then there must be a danger." *Journal of Behavior Therapy and Experimental Psychiatry, 43*, 1032–1038. <http://dx.doi.org/10.1016/j.jbtep.2012.04.005>
- Gao, F., & Chen, L. (2005). Bayesian or non-Bayesian: A comparison study of item parameter estimation in the three-parameter logistic model. *Applied Measurement in Education, 18*, 351–380. http://dx.doi.org/10.1207/s15324818ame1804_2
- Garrard, P., Rentoumi, V., Gesierich, B., Miller, B., & Gorno-Tempini, M. L. (2014). Machine learning approaches to diagnosis and laterality effects in semantic dementia discourse. *Cortex, 55*, 122–129. <http://dx.doi.org/10.1016/j.cortex.2013.05.008>
- Geerlings, H., Laros, J. A., Tellegen, P. J., & Glas, C. A. W. (2014). Testing the difficulty theory of the SON-R 5(1/2)-17, a non-verbal test of intelligence. *British Journal of Mathematical and Statistical Psychology, 67*, 248–265. <http://dx.doi.org/10.1111/bmsp.12017>
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (Vol. 2). London, UK: Chapman & Hall CRC.
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*, 457–472. <http://dx.doi.org/10.1214/ss/1177011136>
- Gelman, A., & Shalizi, C. R. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology, 66*, 8–38. <http://dx.doi.org/10.1111/j.2044-8317.2011.02037.x>
- Geweke, J. (1991). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*. Minneapolis, MN: Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN.
- Gill, J. (2014). *Bayesian methods: A social and behavioral sciences approach* (Vol. 20). Boca Raton, FL: CRC Press.
- Golay, P., Reverte, I., Rossier, J., Favez, N., & Lecerf, T. (2013). Further insights on the French WISC-IV factor structure through Bayesian structural equation modeling. *Psychological Assessment, 25*, 496–508. <http://dx.doi.org/10.1037/a0030676>
- Gong, M., Baron, J., & Kunreuther, H. (2009). Group cooperation under uncertainty. *Journal of Risk and Uncertainty, 39*, 251–270. <http://dx.doi.org/10.1007/s11166-009-9080-2>
- Gopnik, A., Griffiths, T. L., & Lucas, C. G. (2015). When younger learners can be better (or at least more open-minded) than older ones. *Current Directions in Psychological Science, 24*, 87–92. <http://dx.doi.org/10.1177/0963721414556653>
- Haig, B. D. (2009). Inference to the best explanation: A neglected approach to theory appraisal in psychology. *The American Journal of Psychology, 122*, 219–234.
- Hashim, N., & Parris, B. A. (2015). Assessing stimulus-stimulus (semantic) conflict in the Stroop task using saccadic two-to-one color response mapping and prereponse pupillary measures. *Attention, Perception, & Psychophysics, 77*, 2601–2610. <http://dx.doi.org/10.3758/s13414-015-0971-9>
- Heeren, A., Maurage, P., & Philippot, P. (2013). A Bayesian case-controls exploration of the malleability of attentional bias for threat in social phobia. *International Journal of Cognitive Therapy, 6*, 24–39. <http://dx.doi.org/10.1521/ijct.2013.6.1.24>
- Hojtink, H. (2001). Confirmatory latent class analysis: Model selection using Bayes factors and (pseudo) likelihood ratio statistics. *Multivariate Behavioral Research, 36*, 563–588. http://dx.doi.org/10.1207/S15327906MBR3604_04
- Hojtink, H., Béland, S., & Vermeulen, J. A. (2014). Cognitive diagnostic assessment via Bayesian evaluation of informative diagnostic hypotheses. *Psychological Methods, 19*, 21–38. <http://dx.doi.org/10.1037/a0034176>
- Hojtink, H., & Molenaar, I. W. (1997). A multidimensional item response model: Constrained latent class analysis using the Gibbs sampler and posterior predictive checks. *Psychometrika, 62*, 171–189. <http://dx.doi.org/10.1007/BF02295273>
- Hojtink, H., van Kooten, P., & Hulsker, K. (2016a). Bayes factors have frequency properties—This should not be ignored: A rejoinder to Morey, Wagenmakers, and Rouder. *Multivariate Behavioral Research, 51*, 20–22. <http://dx.doi.org/10.1080/00273171.2015.1071705>
- Hojtink, H., van Kooten, P., & Hulsker, K. (2016b). Why Bayesian psychologists should change the way they use the Bayes factor. *Multivariate Behavioral Research, 51*, 2–10. <http://dx.doi.org/10.1080/00273171.2014.969364>
- Hox, J. J., Moerbeek, M., Kluytmans, A., & van de Schoot, R. (2014). Analyzing indirect effects in cluster randomized trials. The effect of estimation method, number of groups and group sizes on accuracy and power. *Frontiers in Psychology, 5*, 78.
- Hox, J., van de Schoot, R., & Matthijsse, S. (2012). How few countries will do? Comparative survey analysis from a Bayesian perspective. *Survey Research Methods, 6*, 87–93.
- Humphries, M. A., Bruno, R., Karpievitch, Y., & Wotherspoon, S. (2015). The expectancy valence model of the Iowa gambling task: Can it produce reliable estimates for individuals? *Journal of Mathematical Psychology, 64–65*, 17–34. <http://dx.doi.org/10.1016/j.jmp.2014.10.002>
- Imel, Z. E., Baldwin, S., Atkins, D. C., Owen, J., Baardseth, T., & Wampold, B. E. (2011). Racial/ethnic disparities in therapist effectiveness: A conceptualization and initial study of cultural competence. *Journal of Counseling Psychology, 58*, 290–298. <http://dx.doi.org/10.1037/a0023284>

- Ip, E. H. (2010). Interpretation of the three-parameter testlet response model and information function. *Applied Psychological Measurement, 34*, 467–482. <http://dx.doi.org/10.1177/0146621610364975>
- Jackson, B., Gucciardi, D. F., & Dimmock, J. A. (2014). Toward a multidimensional model of athletes' commitment to coach-athlete relationships and interdependent sport teams: A substantive-methodological synergy. *Journal of Sport & Exercise Psychology, 36*, 52–68. <http://dx.doi.org/10.1123/jsep.2013-0038>
- Jannarone, R. J., Yu, K. F., & Laughlin, J. E. (1990). Easy Bayes estimation for Rasch-type models. *Psychometrika, 55*, 449–460. <http://dx.doi.org/10.1007/BF02294760>
- Jones, W. P. (2014). Enhancing a short measure of big five personality traits with Bayesian scaling. *Educational and Psychological Measurement, 74*, 1049–1066. <http://dx.doi.org/10.1177/0013164414525040>
- Juslin, P., Nilsson, H., Winman, A., & Lindskog, M. (2011). Reducing cognitive biases in probabilistic reasoning by the use of logarithm formats. *Cognition, 120*, 248–267. <http://dx.doi.org/10.1016/j.cognition.2011.05.004>
- Kaplan, D. (2014). *Bayesian statistics for the social sciences*. New York, NY: Guilford Press Publications.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*, 773–795. <http://dx.doi.org/10.1080/01621459.1995.10476572>
- Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review, 116*, 20–58. <http://dx.doi.org/10.1037/a0014282>
- Kennedy, J. E. (2014). Bayesian and classical hypothesis testing: Practical differences for a controversial area of research. *Journal of Parapsychology, 78*, 170–182.
- Kieftenbeld, V., & Natesan, P. (2012). Recovery of graded response model parameters: A comparison of marginal maximum likelihood and Markov Chain Monte Carlo Estimation. *Applied Psychological Measurement, 36*, 399–419. <http://dx.doi.org/10.1177/0146621612446170>
- Kinoshita, S., & Norris, D. (2012). Task-dependent masked priming effects in visual word recognition. *Frontiers in Psychology, 3*, 178.
- Klugkist, I., Laudy, O., & Hoijtink, H. (2005). Inequality constrained analysis of variance: A Bayesian approach. *Psychological Methods, 10*, 477–493. <http://dx.doi.org/10.1037/1082-989X.10.4.477>
- Klugkist, I., Van Wesel, F., & Bullens, J. (2011). Do we know what we test and do we test what we want to know? *International Journal of Behavioral Development, 35*, 550–560. <http://dx.doi.org/10.1177/0165025411425873>
- Knops, A., Zitzmann, S., & McCrink, K. (2013). Examining the presence and determinants of operational momentum in childhood. *Frontiers in Psychology, 4*, 325.
- Konijn, E. A., van de Schoot, R., Winter, S. D., & Ferguson, C. J. (2015). Possible solution to publication bias through Bayesian statistics, including proper null hypothesis testing. *Communication Methods and Measures, 9*, 280–302. <http://dx.doi.org/10.1080/19312458.2015.1096332>
- Krivitsky, P. N., Handcock, M. S., Raftery, A. E., & Hoff, P. D. (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks, 31*, 204–213. <http://dx.doi.org/10.1016/j.socnet.2009.04.001>
- Kruschke, J. K. (2010). Bayesian data analysis. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*, 658–676. <http://dx.doi.org/10.1002/wcs.72>
- Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science, 6*, 299–312. <http://dx.doi.org/10.1177/1745691611406925>
- Kruschke, J. K. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. New York, NY: Academic Press.
- Kunreuther, H., Silvasi, G., Bradlow, E. T., & Small, D. (2009). Bayesian analysis of deterministic and stochastic prisoner's dilemma games. *Judgment and Decision Making, 4*, 363–384.
- Laplace, P. S. (1814). *Essai Philosophique sur les Probabilités* [A philosophical essay on probabilities]. New York, NY: Courier.
- Lee, J., Choi, J. Y., & Cho, Y. (2011). A forecast simulation analysis of the next-generation DVD market based on consumer preference data. *International Journal of Consumer Studies, 35*, 448–457. <http://dx.doi.org/10.1111/j.1470-6431.2010.00958.x>
- Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology, 55*, 1–7. <http://dx.doi.org/10.1016/j.jmp.2010.08.013>
- Lee, M. D., & Sarnecka, B. W. (2011). Number-knower levels in young children: Insights from Bayesian modeling. *Cognition, 120*, 391–402. <http://dx.doi.org/10.1016/j.cognition.2010.10.003>
- Lee, S. Y., Song, X. Y., & Tang, N. S. (2007). Bayesian methods for analyzing structural equation models with covariates, interaction, and quadratic latent variables. *Structural Equation Modeling, 14*, 404–434. <http://dx.doi.org/10.1080/10705510701301511>
- Lee, S. Y., & Zhu, H. T. (2000). Statistical analysis of nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology, 53*, 209–232. <http://dx.doi.org/10.1348/000711000159303>
- Liddell, T. M., & Kruschke, J. K. (2014). Ostracism and fines in a public goods game with accidental contributions: The importance of punishment type. *Judgment and Decision Making, 9*, 523–547.
- Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., . . . Wagenmakers, E.-J. (2015). JASP (Version 0.7) [Computer software]. Amsterdam, the Netherlands: JASP Project. Retrieved from <https://jasp-stats.org>
- Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). *The BUGS book: A practical introduction to Bayesian analysis*. Boca Raton, FL: CRC Press.
- Lunn, D., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS—a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing, 10*, 325–337. <http://dx.doi.org/10.1023/A:1008929526011>
- Lynch, S. M. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists*. New York, NY: Springer Science & Business Media. <http://dx.doi.org/10.1007/978-0-387-71265-9>
- Lynch, S. M., Brown, J. S., & Harmsen, K. G. (2003). The effect of altering ADL thresholds on active life expectancy estimates for older persons. *The Journals of Gerontology Series B, Psychological Sciences and Social Sciences, 58*, S171–S178. <http://dx.doi.org/10.1093/geronb/58.3.S171>
- Matzke, D., Nieuwenhuis, S., van Rijn, H., Slagter, H. A., van der Molen, M. W., & Wagenmakers, E. J. (2015). The effect of horizontal eye movements on free recall: A preregistered adversarial collaboration. *Journal of Experimental Psychology: General, 144*, e1–e15. <http://dx.doi.org/10.1037/xge0000038>
- McFall, R. M., & Treat, T. A. (1999). Quantifying the information value of clinical assessments with signal detection theory. *Annual Review of Psychology, 50*, 215–241. <http://dx.doi.org/10.1146/annurev.psych.50.1.215>
- Merkle, E., & Rosseel, Y. (2015). *blavaan: Bayesian Latent Variable Analysis* (Version R package version 0.1–1). Retrieved from <http://CRAN.R-project.org/package=blavaan>
- Milfont, T. L., & Sibley, C. G. (2014). The hierarchy enforcement hypothesis of environmental exploitation: A social dominance perspective. *Journal of Experimental Social Psychology, 55*, 188–193. <http://dx.doi.org/10.1016/j.jesp.2014.07.006>
- Montague, M., Krawec, J., Enders, C., & Dietz, S. (2014). The effects of cognitive strategy instruction on math problem solving of middle-school students of varying ability. *Journal of Educational Psychology, 106*, 469–481. <http://dx.doi.org/10.1037/a0035176>
- Moore, T. M., Reise, S. P., Depaoli, S., & Haviland, M. G. (2015). Iteration of partially specified target matrices: Applications in exploratory and

- Bayesian confirmatory factor analysis. *Multivariate Behavioral Research*, 50, 149–161. <http://dx.doi.org/10.1080/00273171.2014.973990>
- Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods*, 16, 406–419. <http://dx.doi.org/10.1037/a0024377>
- Morey, R. D., Wagenmakers, E.-J., & Rouder, J. N. (in press). “Calibrated” Bayes factors should not be used: A reply to Hoijtink, van Kooten, and Hulsker. *Multivariate Behavioral Research*.
- Morrell, S., Taylor, R., Quine, S., Kerr, C., & Western, J. (1994). A cohort study of unemployment as a cause of psychological disturbance in Australian youth. *Social Science & Medicine*, 38, 1553–1564. [http://dx.doi.org/10.1016/0277-9536\(94\)90117-1](http://dx.doi.org/10.1016/0277-9536(94)90117-1)
- Mulder, J., Klugkist, I., van de Schoot, R., Meeus, W. H. J., Selfhout, M., & Hoijtink, H. (2009). Bayesian model selection of informative hypotheses for repeated measurements. *Journal of Mathematical Psychology*, 53, 530–546. <http://dx.doi.org/10.1016/j.jmp.2009.09.003>
- Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, 17, 313–335. <http://dx.doi.org/10.1037/a0026802>
- Muthén, L. K., & Muthén, B. O. (1998–2015). *Mplus user's guide*, 7th edition. Los Angeles, CA: Author.
- Natesan, P. (2015). Comparing interval estimates for small sample ordinal CFA models. *Frontiers in Psychology*, 6, 1599.
- Navarro, D. J., & Perfors, A. F. (2010). Similarity, feature discovery, and the size principle. *Acta Psychologica*, 133, 256–268. <http://dx.doi.org/10.1016/j.actpsy.2009.10.008>
- Neyens, D. M., Boyle, L. N., & Schultheis, M. T. (2015). The Effects of driver distraction for individuals with traumatic brain injuries. *Human Factors*, 57, 1472–1488. <http://dx.doi.org/10.1177/0018720815594057>
- O'Donnell, B. M., & Colombo, E. M. (2011). Foveal spatial summation in human cone mechanism. *Perception*, 40, 50–62. <http://dx.doi.org/10.1068/p6775>
- Owen, J., Drinane, J., Tao, K. W., Adelson, J. L., Hook, J. N., Davis, D., & Fookune, N. (2015). Racial/ethnic disparities in client unilateral termination: The role of therapists' cultural comfort. *Psychotherapy Research*. Advance online publication. <http://dx.doi.org/10.1080/10503307.2015.1078517>
- Ozechowski, T. J. (2014). Empirical Bayes MCMC estimation for modeling treatment processes, mechanisms of change, and clinical outcomes in small samples. *Journal of Consulting and Clinical Psychology*, 82, 854–867. <http://dx.doi.org/10.1037/a0035889>
- Pearl, J. (1985). *Bayesian networks: A model of self-activated memory for evidential reasoning*. (UCLA Tech. Rep. No. CSD-850017). Retrieved from http://ftp.cs.ucla.edu/tech-report/198_-reports/850017.pdf
- Plummer, M. (2016). *Rjags: Bayesian Graphical Models using MCMC. R package, version 4–6*. Retrieved from <https://CRAN.R-project.org/package=rjags>
- Pullenayegum, E. M., Guo, Q., & Hopkins, R. B. (2012). Developing critical thinking about reporting of Bayesian analyses. *Journal of Statistics Education*, 20(1), n1.
- Rietbergen, C. (2016). *Quantitative Evidence Synthesis with Power Priors* (Doctoral dissertation). Retrieved from ISBN. (978-90-6464-971-4). <http://dspace.library.uu.nl/handle/1874/329030>
- Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, 44, 1276–1284. <http://dx.doi.org/10.1037/0003-066x.44.10.1276>
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56, 356–374. <http://dx.doi.org/10.1016/j.jmp.2012.08.001>
- Rouder, J. N., Sun, D., Speckman, P. L., Lu, J., & Zhou, D. (2003). A hierarchical Bayesian statistical framework for response time distributions. *Psychometrika*, 68, 589–606. <http://dx.doi.org/10.1007/BF02295614>
- Rupp, A. A., Dey, D. K., & Zumbo, B. D. (2004). To Bayes or not to Bayes, from whether to when: Applications of Bayesian methodology to modeling. *Structural Equation Modeling*, 11, 424–451. http://dx.doi.org/10.1207/s15328007sem1103_7
- Ryoo, J. H., Molfese, V. J., Brown, E. T., Karp, K. S., Welch, G. W., & Bovaird, J. A. (2015). Examining factor structures on the Test of Early Mathematics Ability–3: A longitudinal approach. *Learning and Individual Differences*, 41, 21–29. <http://dx.doi.org/10.1016/j.lindif.2015.06.003>
- Scheibehenne, B., & Studer, B. (2014). A hierarchical Bayesian model of the influence of run length on sequential predictions. *Psychonomic Bulletin & Review*, 21, 211–217. <http://dx.doi.org/10.3758/s13423-013-0469-1>
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.
- Serra, G., Koukopoulos, A., De Chiara, L., Napoletano, F., Koukopoulos, A. E., Curto, M., . . . Baldessarini, R. J. (2015). Features preceding diagnosis of bipolar versus major depressive disorders. *Journal of Affective Disorders*, 173, 134–142. <http://dx.doi.org/10.1016/j.jad.2014.10.050>
- Sims, C. R., Neth, H., Jacobs, R. A., & Gray, W. D. (2013). Melioration as rational choice: Sequential decision making in uncertain environments. *Psychological Review*, 120, 139–154. <http://dx.doi.org/10.1037/a0030850>
- Sinharay, S. (2005). Assessing fit of unidimensional item response theory models using a Bayesian approach. *Journal of Educational Measurement*, 42, 375–394. <http://dx.doi.org/10.1111/j.1745-3984.2005.00021.x>
- Smith, M., Kohn, R., & Mathur, S. K. (2000). Bayesian semiparametric regression: An exposition and application to print advertising data. *Journal of Business Research*, 49, 229–243. [http://dx.doi.org/10.1016/S0148-2963\(99\)00055-7](http://dx.doi.org/10.1016/S0148-2963(99)00055-7)
- Sohlberg, S., & Andersson, G. (2005). Extracting a maximum of useful information from statistical research data. *Scandinavian Journal of Psychology*, 46, 69–77. <http://dx.doi.org/10.1111/j.1467-9450.2005.00436.x>
- Song, X. Y., & Lee, S. Y. (2002). Analysis of structural equation model with ignorable missing continuous and polytomous data. *Psychometrika*, 67, 261–288. <http://dx.doi.org/10.1007/BF02294846>
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B. Methodological*, 64, 583–639. <http://dx.doi.org/10.1111/1467-9868.00353>
- Spiegelhalter, D. J., Myles, J. P., Jones, D. R., & Abrams, K. R. (2000). Bayesian methods in health technology assessment: A review. *Health Technology Assessment*, 4, 1–130.
- Stern, H. S. (2016). A test by any other name: P values, Bayes factors, and statistical inference. *Multivariate Behavioral Research*, 51, 23–29. <http://dx.doi.org/10.1080/00273171.2015.1099032>
- Steyvers, M., Lee, M. D., & Wagenmakers, E. J. (2009). A Bayesian analysis of human decision-making on bandit problems. *Journal of Mathematical Psychology*, 53, 168–179. <http://dx.doi.org/10.1016/j.jmp.2008.11.002>
- Sung, L., Hayden, J., Greenberg, M. L., Koren, G., Feldman, B. M., & Tomlinson, G. A. (2005). Seven items were identified for inclusion when reporting a Bayesian analysis of a clinical study. *Journal of Clinical Epidemiology*, 58, 261–268. <http://dx.doi.org/10.1016/j.jclinepi.2004.08.010>
- Tondo, L., Visioli, C., Preti, A., & Baldessarini, R. J. (2014). Bipolar disorders following initial depression: Modeling predictive clinical factors. *Journal of Affective Disorders*, 167, 44–49. <http://dx.doi.org/10.1016/j.jad.2014.05.043>
- Tong, J., Mao, O., & Goldreich, D. (2013). Two-point orientation discrimination versus the traditional two-point test for tactile spatial acuity assessment. *Frontiers in Human Neuroscience*, 7, 579.

- Tsutakawa, R. K., & Johnson, J. C. (1990). The effect of uncertainty of item parameter estimation on ability estimates. *Psychometrika*, *55*, 371–390. <http://dx.doi.org/10.1007/BF02295293>
- Van Den Hout, M., Gangemi, A., Mancini, F., Engelhard, I. M., Rijkeboer, M. M., Van Dams, M., & Klugkist, I. (2014). Behavior as information about threat in anxiety disorders: A comparison of patients with anxiety disorders and non-anxious controls. *Journal of behavior therapy and experimental psychiatry*, *45*, 489–495.
- van de Schoot, R., Broere, J. J., Perryck, K. H., Zondervan-Zwijenburg, M., & van Loey, N. E. (2015). Analyzing small data sets using Bayesian estimation: The case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *European Journal of Psychotraumatology*, *6*, 25216. <http://dx.doi.org/10.3402/ejpt.v6.25216>
- van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*, *85*, 842–860. <http://dx.doi.org/10.1111/cdev.12169>
- van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthén, B. (2013). Facing off with Scylla and Charybdis: A comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology*, *4*, 770. <http://dx.doi.org/10.3389/fpsyg.2013.00770>
- van de Schoot, R., Mulder, J., Hoijsink, H., van Aken, M. A. G., Dubas, J. S., de Castro, B. O., . . . Romeijn, J. W. (2011). An introduction to Bayesian model selection for evaluating informative hypotheses. *European Journal of Developmental Psychology*, *8*, 713–729. <http://dx.doi.org/10.1080/17405629.2011.621799>
- van Ravenzwaaij, D., Dutilh, G., & Wagenmakers, E. J. (2011). Cognitive model decomposition of the BART: Assessment and application. *Journal of Mathematical Psychology*, *55*, 94–105. <http://dx.doi.org/10.1016/j.jmp.2010.08.010>
- van Well, S., Kolk, A. M., & Klugkist, I. G. (2008). Effects of sex, gender role identification, and gender relevance of two types of stressors on cardiovascular and subjective responses: Sex and gender match and mismatch effects. *Behavior Modification*, *32*, 427–449. <http://dx.doi.org/10.1177/0145445507309030>
- Verguts, T., & De Boeck, P. (2000). A Rasch model for detecting learning while solving an intelligence test. *Applied Psychological Measurement*, *24*, 151–162. <http://dx.doi.org/10.1177/01466210022031589>
- Voorspoels, W., Navarro, D. J., Perfors, A., Ransom, K., & Storms, G. (2015). How do people learn from negative evidence? Non-monotonic generalizations and sampling assumptions in inductive reasoning. *Cognitive Psychology*, *81*, 1–25. <http://dx.doi.org/10.1016/j.cogpsych.2015.07.001>
- Voorspoels, W., Storms, G., & Vanpaemel, W. (2012). Contrast effects in typicality judgements: A hierarchical Bayesian approach. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *65*, 1721–1739. <http://dx.doi.org/10.1080/17470218.2012.662237>
- Wagenmakers, E. J., Lee, M., Lodewyckx, T., & Iverson, G. J. (2008). Bayesian versus frequentist inference. In H. Hoijsink, I. Klugkist, & P. A. Boelen (Eds.), *Bayesian evaluation of informative hypotheses* (pp. 181–207). New York, NY: Springer New York. http://dx.doi.org/10.1007/978-0-387-09612-4_9
- Wagenmakers, E. J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognitive Psychology*, *60*, 158–189. <http://dx.doi.org/10.1016/j.cogpsych.2009.12.001>
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. L. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, *100*, 426–432. <http://dx.doi.org/10.1037/a0022790>
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*, 632–638. <http://dx.doi.org/10.1177/1745691612463078>
- Wainer, H., Wang, X. A., Skorupski, W. P., & Bradlow, E. T. (2005). A Bayesian method for evaluating passing scores: The PPoP curve. *Journal of Educational Measurement*, *42*, 271–281. <http://dx.doi.org/10.1111/j.1745-3984.2005.00014.x>
- Wang, C., & Nydick, S. W. (2015). Comparing two algorithms for calibrating the restricted non-compensatory multidimensional IRT model. *Applied Psychological Measurement*, *39*, 119–134. <http://dx.doi.org/10.1177/0146621614545983>
- Wang, W. C., Chen, H. F., & Jin, K. Y. (2015). Item response theory models for wording effects in mixed-format scales. *Educational and Psychological Measurement*, *75*, 157–178. <http://dx.doi.org/10.1177/0013164414528209>
- Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology*, *44*, 92–107. <http://dx.doi.org/10.1006/jmps.1999.1278>
- Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science*, *6*, 291–298. <http://dx.doi.org/10.1177/1745691611406923>
- Wilson, S. J., Barrineau, M. J., Butner, J., & Berg, C. A. (2014). Shared possible selves, other-focus, and perceived wellness of couples with prostate cancer. *Journal of Family Psychology*, *28*, 684–691. <http://dx.doi.org/10.1037/fam0000015>
- Wollack, J. A., Bolt, D. M., Cohen, A. S., & Lee, Y. S. (2002). Recovery of item parameters in the nominal response model: A comparison of marginal maximum likelihood estimation and Markov chain Monte Carlo estimation. *Applied Psychological Measurement*, *26*, 339–352. <http://dx.doi.org/10.1177/0146621602026003007>
- Wong, T. M. L., & Van de Schoot, R. (2012). The effect of offenders' sex on reporting crimes to the police. *Journal of Interpersonal Violence*, *27*, 1276–1292. <http://dx.doi.org/10.1177/0886260511425245>
- Yu, M. S., & Huang, F. L. (2003). Disambiguating the senses of non-text symbols for Mandarin TTS systems with a three-layer classifier. *Speech Communication*, *39*, 191–229. [http://dx.doi.org/10.1016/S0167-6393\(02\)00015-8](http://dx.doi.org/10.1016/S0167-6393(02)00015-8)
- Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods*, *14*, 301–322. <http://dx.doi.org/10.1037/a0016972>
- Zhang, D., Liang, B., Wu, X., Wang, Z., Xu, P., Chang, S., . . . Huang, R. (2015). Directionality of large-scale resting-state brain networks during eyes open and eyes closed conditions. *Frontiers in Human Neuroscience*, *9*, 81.

Received August 18, 2015

Revision received June 2, 2016

Accepted June 19, 2016 ■