

A Bayesian Latent Group Analysis for Detecting Poor Effort in the Assessment of Malingering

Alonso Ortega^{1,2,*}, Eric-Jan Wagenmakers³, Michael D. Lee⁴, Hans J. Markowitsch^{1,5}, Martina Piefke^{1,5}

¹*Department of Physiological Psychology, Bielefeld University, Bielefeld, Germany*

²*Escuela de Psicología, Facultad de Medicina, Universidad de Valparaíso, Valparaíso, Chile*

³*Department of Psychological Methods, University of Amsterdam, Amsterdam, The Netherlands*

⁴*Department of Cognitive Sciences, University of California, Irvine, CA, USA*

⁵*Center of Excellence “Cognitive Interaction Technology” (CITEC), Bielefeld University, Bielefeld, Germany*

*Corresponding author at: Department of Physiological Psychology, Bielefeld University, Universitätsstr. 25, Bielefeld 33615, Germany.

Tel.: +49-521-106-44-82; fax: +49-521-106-60-49.

E-mail address: alonso.ortega@uv.cl (A. Ortega).

Accepted 13 March 2012

Abstract

Despite their theoretical appeal, Bayesian methods for the assessment of poor effort and malingering are still rarely used in neuropsychological research and clinical diagnosis. In this article, we outline a novel and easy-to-use Bayesian latent group analysis of malingering whose goal is to identify participants displaying poor effort when tested. Our Bayesian approach also quantifies the confidence with which each participant is classified and estimates the base rates of malingering from the observed data. We implement our Bayesian approach and compare its utility in effort assessment to that of the classic below-chance criterion of symptom validity testing (SVT). In two experiments, we evaluate the accuracy of both a Bayesian latent group analysis and the below-chance criterion of SVT in recovering the membership of participants assigned to the malingering group. Experiment 1 uses a simulation research design, whereas Experiment 2 involves the differentiation of patients with a history of stroke from coached malingerers. In both experiments, sensitivity levels are high for the Bayesian method, but low for the below-chance criterion of SVT. Additionally, the Bayesian approach proves to be resistant to possible effects of coaching. We conclude that Bayesian latent group methods complement existing methods in making more informed choices about malingering.

Keywords: Malingering; Symptom validity testing; Poor effort; Coaching; Bayesian methods; Hierarchical models

Introduction

In order to obtain valid results, neuropsychological tests generally require that the examinee exert full effort (Larrabee, 2007). Indeed, poor effort may impact more on test results than neurological conditions or brain injury (Denney, 2008). The term “effort” refers to a subject’s motivation to perform well during psychological assessment, and poor effort is usually the process that underlies malingering (Iverson, 2003). The Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) defines malingering as “the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives such as avoiding military duty, avoiding work, obtaining financial compensation, evading criminal prosecution, or obtaining drugs” (American Psychiatric Association [DSM-IV-TR], 2000, p. 739).

Slick, Sherman, and Iverson (1999) proposed some criteria for the determination of malingering. These criteria require evidence from especially developed effort measures, among other indicators (see also Boone, 2007; Larrabee, Greiffenstein, Grewe, & Bianchini, 2007). Nevertheless, every diagnostic evaluation process entails the risk of misclassification, the consequences of which may bring serious legal, economic, or clinical repercussions. On the one hand, to misclassify malingerers as patients with genuine cognitive impairment risks providing people with undeserved financial or legal benefits. On the other hand, it is especially grievous to misclassify patients with genuine cognitive impairment as malingerers, not only denying

them the benefits that they deserve, but also unjustly labeling them as frauds. These arguments support and mandate the permanent revision and improvement of effort measures and other malingering detection strategies.

During the last decade, the inclusion of effort measures in neuropsychological assessment gained growing interest among researchers and practitioners (Bush et al., 2005; Green, 2003; Horton, 2010). One of the most widely used measures is symptom validity testing (SVT; Grote & Hook, 2007). Nevertheless, as will be described later, SVT has been criticized for its low sensitivity.

The main goal of this article is to introduce a Bayesian latent group analysis as a novel procedure for effort assessment in the context of malingering detection. Bayesian methods are popular in fields such as statistics and machine learning (Poirier, 2006), and they are rapidly gaining grounds in more applied fields such as genetics (Stephens & Balding, 2009). Within the field of psychology, Bayesian approaches are widely used in psychophysics (e.g., Clemens, De Vrijer, Selen, Van Gisbergen, & Medendorp, 2011) and form the basis of some methods for data analysis in functional neuroimaging research on memory and other cognitive functions (e.g., Piefke, Onur, & Fink, 2010; Piefke, Weiss, Zilles, Markowitsch, & Fink, 2003; Schulte-Rüther, Markowitsch, Fink, & Piefke, 2007), as well as memory malingering (e.g., Larsen, Allen, Bigler, Goodrich-Hunsaker, & Hopkins, 2010; Lee et al., 2009; Wolfe et al., 2010). Moreover, the Bayesian approach allows researchers and clinicians to quantify the certainty of their judgment for each individual person. If desired, the Bayesian approach can be extended to incorporate utilities, and so make decisions sensitive to the costs and benefits associated with each classification decision.

Symptom Validity Testing

SVT is based on the forced-choice method that was originally developed in the context of psychophysics (Macmillan & Creelman, 2008). In the mid-seventies, Pankratz, Fausti, and Peed (1975) implemented this technique to distinguish deafness or hysterical loss of hearing (i.e., conversion disorder) from malingering in a young patient who made repeated attempts to obtain compensation for his disabilities. Results of the study promoted the use of the forced-choice testing procedure as a reliable method to assess both sensory deficits and malingering.

The rationale underlying SVT is simple. Given a particular number of forced-choice trials, the probability of obtaining a determined number of successes by chance alone is estimated using the normal approximation to the binomial distribution. Then, scores that fall significantly below the chance level “would seem to offer rather definitive evidence of an intentional (i.e., non-chance) attempt to perform poorly on the test by active avoidance of the correct response” (Grote & Hook, 2007, p. 45). This inference is plausible because most SVT are relatively easy, even for severe cognitive impaired patients (Morel & Shepherd, 2008; Slick et al., 2003).

The Sensitivity Problem of SVT

For many years, the standard approach in malingering assessment has been to achieve high specificity at the expense of sensitivity, in order to reduce the false-positive rates (Iverson, 2007). On the one hand, this turns the below-chance criterion of SVT into a safe and reliable index of poor effort when results are positives. On the other hand, the inherent conservatism of the approach only allows for the identification of a small subset of malingerers (Beetar & Williams, 1995; Rogers, 2008; Slick et al., 2003).

In order to increase the low sensitivity of SVTs, researchers working on test construction began to derive empirically cutoff scores (Iverson & Binder, 2000). Currently, many SVTs (e.g., the Portland Digit Recognition Test [PDRT], Binder, 1993; the Test of Memory Malingering [TOMM], Tombaugh, 1996; the Validity Indicator Profile [VIP], Frederick, 1997; and the Word Memory Test [WMT], Green, 2003) do not use the below-chance criterion as the primary decision rule for the determination of poor effort (Frederick & Speed, 2007). Although the use of empirically derived cutoff scores has some advantages, such as increasing sensitivity levels and diminishing false-negative rates, it may also introduce some difficulties. First, when a test is used in different populations or settings, specific norms need to be obtained (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Second, sensitivity and specificity are fixed properties of a test only as long as this is used with similar groups of people (Streiner, 2003). As a consequence, these indices have also to be recalculated when the test is used in populations with different characteristics. Third, further revisions must be made when the validity of test score interpretations is compromised by new research data, substantial changes in the domain of application, or new conditions of test use are recommended (American Educational Research Association et al., 1999). These arguments suggest that test norms and accuracy classification indices should be updated periodically, which generally entails a considerable expenditure of time, resources, and effort. In sum, the use of empirically

derived norms improved SVT's sensitivity but also undermined its original purpose and main advantage that is to determine poor effort by simply using the below-chance criterion of SVT.

Being aware of this problem, some researchers proposed the use of statistical measures such as odds and likelihood ratios, or Bayesian methods as strategies for effort assessment (Iverson, 2007; Millis, 2008). Here, we proposed a Bayesian method because it yields intuitive results and because it takes into account prior knowledge (or lack thereof) about malingering base rates and about commonalities and differences between individual participants. In addition, Bayesian methods do not necessarily require complex and time-consuming methods for the construction and validation of malingering instruments (Mossman, 2000). Finally, Bayesian methods allow automatic updating of knowledge as more information becomes available, such that information from different groups of people can be added to the existing database in a hierarchical or a multilevel framework.

Bayesian Analysis for Effort Assessment: An Overview

A Bayesian analysis generally involves the updating of prior information in light of newly available experimental data (Samaniego, 2010). This method combines three sources of information: (a) a model that says how latent parameters generate data, (b) the prior probability distribution, which represents advance knowledge about the parameters, and (c) the observed data. Following the Bayes theorem, the combination of these three elements produces the posterior probability distribution, which represents the knowledge about the model parameters after the data have been observed. The change from the prior to the posterior distribution reflects what has been learned from the data. Posterior distributions that are relatively peaked indicate relatively precise knowledge, whereas posterior distributions that are relatively flat indicate that knowledge is vague or uncertain. For more detailed information about Bayesian inference, see, for instance, O'Hagan and Forster (2004), Kruschke and John (2010), Lee and Wagenmakers (2010), and Dienes (2011).

Mossman and Hart (1996) first proposed the use of Bayes' theorem for estimating the level of effort displayed by an examinee. In their study, Mossman and Hart (1996) re-analyzed data from previously published malingering studies. Results allowed for a precise estimation of the probability that an examinee is feigning some cognitive or emotional impairment (Mossman & Hart, 1996). However, the use of the Bayesian approach for effort testing was discouraged soon after it was first proposed. The main reason for this was the existence of high variability in the base rates of malingering. Rogers and Salekin (1998) argued that an implicit assumption of Bayes' theorem is that the base rate of a specific condition (e.g., malingering) is measurable and relatively stable. The available evidence showed, however, that malingering base rates were highly variable (Rogers, Sewell, & Goldstein, 1994; Rogers, Salekin, Sewell, Goldstein, & Leonard, 1998). Based on this Rogers and Salekin (1998) suggested that a Bayesian approach had little clinical utility in assessing effort in the context of malingering.

In recent years, the increase in malingering research has yielded key information about malingering base rates (Larrabee, 2007). A survey study of neuropsychologists involved in forensic work showed a base rate of 30% in persons alleging personal injury, disability, and chronic pain (Mittenberg, Patton, Canyock, & Condit, 2002). The same study also found that the highest base rate of malingering was 38.5% in litigants alleging mild head injury (Mittenberg et al., 2002). Another review encompassing 1,363 mild traumatic brain-injury litigants reports a base rate of 40% for malingering (Larrabee, 2003). Miller, Boyd, Cohn, Wilson, and McFarland (2006) found that 54% of Social Security disability applicants failed in two effort measures. Later, Larrabee (2007) estimated that base rates of malingering approach or exceed 50% for a range of civil and criminal settings.

The updated knowledge about malingering base rates shows values that are subject to less variability than surmised before. This knowledge allows us to propose more specific and informative priors on malingering base rates than a decade ago. In addition, the emergence of modern Bayesian methods for making inferences with more complicated hierarchical models allows the possibility of estimating malingering base rates from the observed data. Together, these empirical and methodological advances address the initial concerns about the appropriateness of Bayesian methods raised by Rogers and Salekin (1998).

General Method

A Bayesian Latent Group Analysis of Malingering

Model specification. The goal of our Bayesian latent group analysis is to identify participants who are displaying poor effort when tested. This Bayesian approach also quantifies the confidence with which each participant is classified and estimates base rates of malingering from the observed data.

The model specification is based on the following assumptions. First, our model assumes the existence of two latent groups: the bona fide group and the malingering group. The former group corresponds to participants who answer the test giving their

best and the latter group corresponds to participants who are instructed to feign cognitive impairment during testing. Second, the model assumes that the bona fide group has an unknown mean rate μ_{bon} of answering any particular test item correctly. The mean rate μ_{bon} is higher than chance (i.e., $.5 < \mu_{\text{bon}} < 1$). A third assumption of the model is that the malingering group has an unknown mean rate μ_{mal} of answering any question correctly and that μ_{bon} is greater than μ_{mal} (or, equivalently, that $\mu_{\text{mal}} = \mu_{\text{bon}} - \mu_{\text{diff}}$, with $\mu_{\text{diff}} > 0$). Together, these assumptions simply state that bona fide participants have a higher success rate than malingerers.

Fourth, our Bayesian approach assumes that participants within each group are similar to each other, but not identical. This psychologically plausible assumption is implemented through a hierarchical structure, where individual parameter estimates θ_i are constrained by group-level distributions (Nilsson et al., 2010; Shiffrin, Lee, Kim, & Wagenmakers, 2008; see also the *Journal of Mathematical Psychology* special issue on hierarchical Bayesian models, Lee, 2010). We use a standard “Beta-binomial” hierarchical model, in which each individual’s success rate is constrained by a group-level Beta distribution. We use the parameterization in which the α and β parameters from the Beta(α , β) distribution are transformed into a group mean $\mu = \alpha/(\alpha + \beta)$ and a group precision $\lambda = \alpha + \beta$. For participants in the bona fide group, individual success rates are assumed to be governed by a group-level Beta distribution with mean μ_{bon} and precision λ_{bon} ; for participants in the malingering group, the group-level Beta distribution has mean μ_{mal} and precision λ_{mal} . Within each group, the similarity of its members is quantified by the precision λ . When λ is high, the group members perform similarly, and when λ is low, the group members perform differently.

Fig. 1 represents the structure of our Bayesian latent group analysis of malingering in graphical model notation. Nodes indicate data or variables, and arrows indicate statistical dependencies. In Fig. 1, nodes n (i.e., the total number of trials) and k_i (i.e., the number of correct answers for each participant i , the maximum of which depends on n) are shaded, indicating

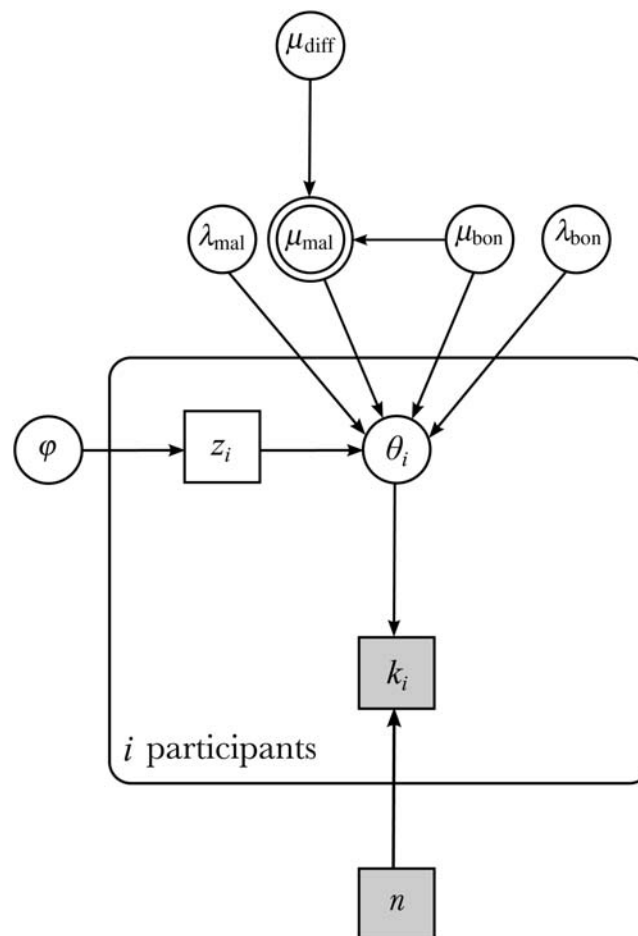


Fig. 1. Graphical model for inferring membership of two latent groups. One group consists of malingerers and the other consists of bona fide participants. Each participant has their own rate of answering test questions correctly, albeit one that is constrained by a group level distribution.

that these variables have been observed. These nodes are also square, indicating that they represent discrete as opposed to continuous values. The plate indicates a structure that is repeated (in this case, it is repeated once for every participant). The node φ represents the prior assumption about the malingering base rate. The binary variable z_i determines group membership; based on z_i , the individual success rate parameter θ_i is constrained either by the group-level distribution for the malingerers (i.e., a Beta distribution with mean μ_{mal} and precision λ_{mal}) or by a group-level distribution for the bona fide participants (i.e., a Beta distribution with mean μ_{bon} and precision λ_{bon}). In the graphical model, μ_{mal} is deterministically calculated as $\text{logit}(\mu_{\text{mal}}) = \text{logit}(\mu_{\text{bon}}) - \mu_{\text{diff}}$, which is why node μ_{mal} has double borders.

Priors on malingering base rates. A Bayesian approach can incorporate different prior information about base rates of malingering and can even estimate these base rates from data. In the graphical model shown in Fig. 1, this information is encoded in the prior for the classification variable z . Specifically, for each subject i , the classification variable has a prior that reflects, or can reflect, advance knowledge about base rates. Thus, z_i has a Bernoulli prior φ , and here, we explore two ways to deal with the uncertainty in this prior.

The first way is to assign φ itself a prior distribution such that it can be estimated from the data. Here, we assigned φ a Beta(5,5) prior, which is relatively uninformative but does not assign a lot of mass to the extremes of the scale. This reflects a prior belief that we are mostly uncertain about the true rate of malingering, but do not believe it to be very low (e.g., 5%) or very high (e.g., 95%).

The second method is essentially a robustness analysis, in which we implemented this Bayesian approach with three different prior beliefs about the base rate of malingering, $\varphi = .5$, $\varphi = .4$, and $\varphi = .3$. These priors cover the plausible range of values for malingering base rates according to the cited literature. If we find that our analyses give essentially the same answers for all these base-rate choices, we can conclude that the data are informative enough that the exact assumptions made about base rates do not matter. On the other hand, if our analyses are different, we have to conclude that the data are more ambiguous, and more work is needed to quantify the prior beliefs for the base rate.

Model implementation. Because the posterior distributions in our Bayesian approach are analytically intractable, we implemented the model using Markov chain Monte Carlo (MCMC; e.g., Gamerman and Lopes, 2006; Gilks, Richardson, & Spiegelhalter, 1996). MCMC sampling is a numerical method that repeatedly draws values from a posterior distribution in order to approximate it to any desired degree of accuracy. This advantage of Bayesian analysis allows us to obtain accurate estimates of posterior probability distributions.

In our studies, the MCMC sampling was accomplished easily in the popular WinBUGS software program (Lunn, Spiegelhalter, Thomas, & Best, 2009; Lunn, Thomas, Best, & Spiegelhalter, 2000). The Technical Appendix (Supplementary material online) provides the required information to implement the Bayesian latent group analysis of malingering. The WinBUGS program uses the code and the data provided in the Technical Appendix (Supplementary material online), conducts the MCMC sampling, and outputs the results (i.e., approximations to the posterior distributions). Using the code provided, interested readers can replicate our results and also analyze their own data sets.

Advantages and novelty. There are four main interesting features of this Bayesian approach. The first one is that the model learns from participants' answers and, in this way, estimates base rates from the data. This constitutes an important advantage and addresses the base rate problem raised by Rogers and Salekin (1998). The second one is the use of a hierarchical model, which has the advantage of being able to account simultaneously for both differences and similarities between participants (e.g., Nilsson, Rieskamp, & Wagenmakers, 2010). Since individual parameters θ_i originate from a group-level distribution, participants are not treated as if their answers were completely independent from the others. This group-level constraint helps avoid the potentially unreliable estimation of a particular individual's parameter by borrowing strength of the information that is obtained from the other individuals (Nilsson et al., 2010). This makes the individual parameter estimates more reliable, because it incorporates more information into their inference. To the best of our knowledge, this approach has not been used before to help in malingering detection. A third advantage is that Bayesian methods apply equally validly to any sample size no matter how small (e.g., Jaynes, 2003; Jeffreys, 1961). This makes Bayesian analysis especially well suited in research fields where participants need to meet very specific characteristics and, therefore, are rather scarce. Finally, the Bayesian model allows one to quantify the certainty or confidence of the classification by means of a probability; for each subject i , this certainty is given by the posterior mean of the classification variable z_i . For instance, suppose the posterior mean of z_3 equals .01, and that of z_{12} equals .43; this indicates that there is a 1% chance that Participant 3 belongs to the group of malingerers, and a 43% chance that Participant 12 belongs to the group of malingerers. Even though one may classify both participants as bona fide, the classification variable indicates that Participant 12 may need to undergo additional testing.

Experiment 1: Bona Fide Versus Naïve Malingerers

Method

To demonstrate the possibilities of the Bayesian approach, we applied the method in two experiments. Experiment 1 used a simulation research design in which participants were randomly assigned to “bona fide” and “naïve malingering” groups. We term the malingering group “naïve” because it was assumed that participants had no previous experience about how to feign cognitive impairment during neuropsychological testing.

We analyze the data using both a Bayesian latent group analysis and the below-chance criterion of the traditional SVT method. The below-chance criterion of the SVT method considered a cutoff score of 17 correct answers as a decision rule to determine the presence of poor effort. This cutoff score was set considering that, from a total of 45 forced-choice trials (two alternatives), a total raw score lower than 17 was expected to occur less than 5% of the time ($p < .05$) by chance alone.

Participants. Twenty-two undergraduate students at Bielefeld University were recruited (age: $M = 24.18$ years; $SD = 5.24$). The students were randomly assigned to a bona fide group ($N = 10$; age: $M = 25.20$ years; $SD = 4.76$) and a naïve malingering group ($N = 12$; age: $M = 23.33$ years; $SD = 5.67$). Written informed consent was obtained from all participants before starting the experiment.

Procedure. Participants assigned to the bona fide group were asked to give their best performance during testing and naïve malingering participants were asked to feign a cognitive impairment. Role instructions to the naïve malingering participants were to “answer the tests as you imagine that a person with cognitive impairment would do.” No further information was provided. Comprehension of role instructions was confirmed before starting the testing. A post-test interview was conducted to corroborate participants’ response style during testing.

Memory recognition forced-choice task. A 45-trial computerized forced-choice task of visual recognition was implemented and applied using MediaLab™ software (Jarvis, 2008). The paradigm consisted of a stimulus presentation phase and a stimulus recognition phase. In the presentation phase, 45 simple colored drawings belonging to four categories (i.e., furniture, food, animals, and clothes) were presented (stimulus onset time = 3 s). Participants were instructed to memorize the presented stimuli. During the recognition phase, each presented drawing (target) was paired with a new drawing (foil). For each pair, participants had to choose the previously presented stimulus.

Neuropsychological battery. Measures of attention, memory, executive functions, and visuospatial abilities were obtained from all participants in order to exclude severe cognitive deficits as a possible confounding variable. None of the participants had to be excluded due to neuropsychological deficits.

Results

Below-chance criterion of SVT. Two participants of the malingering condition did not follow the instructions of faking a cognitive impairment, but rather performed in a normal way and achieved almost perfect scores (44 of 45). Consequently, sensitivity was estimated only considering the outcomes of 10 of 12 participants. Table 1 shows that only 2 of 10 participants of the malingering group fell significantly below the chance level ($z = -2.09$; $p = .05$). This result evidenced the described low sensitivity of the below-chance criterion of SVT. On the other hand, none of the participants of the bona fide group was misclassified. This fact confirmed the high specificity of this method, as shown in Table 1. However, it is important to acknowledge that these results were expected considering that this method allows the achievement of relatively lower sensitivity levels in order to diminish the false-positive rates.

Bayesian latent group analysis of malingering. The classification of participants into the groups in the Bayesian approach is determined from the posterior on the z_i parameters. Table 2 shows the mean posterior value for the classification variable z_i (mean z_i), which are closer to zero for each participant within the bona fide group, indicating a low probability of belonging to the naïve malingering group. Consequently, no participants within the bona fide group were misclassified as malingerer. On the contrary, the mean posterior values for the classification variable z_i are closer to one for each participant within the naïve malingering group. This means that the Bayesian analysis successfully recovered all participants’ group membership and no false positives were found.

Table 1. Individual results of Experiment 1 within the Bona fide and Naïve malingering groups using the below-chance criterion of SVT

Method	Bona fide group			Naïve malingering group		
	Participant _i	Raw score	z-value	Participant _i	Raw score	z-value
SVT's below-chance criterion	P ₁	45	6.559	P ₁₁	30	2.087
	P ₂	45	6.559	P ₁₂	20	-0.596
	P ₃	44	6.261	P ₁₃	6	-4.770*
	P ₄	45	6.559	P ₁₄	44 ^a	6.261
	P ₅	44	6.261	P ₁₅	44 ^a	6.261
	P ₆	45	6.559	P ₁₆	27	1.193
	P ₇	45	6.559	P ₁₇	25	0.596
	P ₈	45	6.559	P ₁₈	17	-1.491
	P ₉	45	6.559	P ₁₉	14	-2.385*
	P ₁₀	45	6.559	P ₂₀	27	1.193
			P ₂₁	35	3.578	
			P ₂₂	30	2.087	

Note: SVT = symptom validity testing.

^aParticipants did not follow the role instructions.

*Participants' scores fell significantly below-chance ($z < -2.09$; $p < .05$).

Table 2. Individual classification probabilities for the Experiment 1 within the Bona fide and Naïve malingering groups assuming different priors on malingering base rates

Posterior individual classification probabilities		Prior on Malingering Base Rates							
		Beta(5,5)		$p(\varphi) = .5$		$p(\varphi) = .4$		$p(\varphi) = .3$	
Group	Participant _i	$p(z_i D)$		$p(z_i D)$		$p(z_i D)$		$p(z_i D)$	
		Mean \bar{z}_i	SD \bar{z}_i	Mean \bar{z}_i	SD \bar{z}_i	Mean \bar{z}_i	SD \bar{z}_i	Mean \bar{z}_i	SD \bar{z}_i
Bona fide	P ₁	.003	.050	.002	.044	.001	.027	.000	.020
	P ₂	.002	.045	.002	.050	.001	.030	.001	.028
	P ₃	.009	.095	.011	.105	.005	.069	.003	.059
	P ₄	.002	.043	.003	.051	.001	.036	.001	.034
	P ₅	.007	.085	.009	.093	.006	.077	.004	.059
	P ₆	.002	.043	.002	.045	.002	.039	.001	.029
	P ₇	.002	.039	.002	.047	.001	.036	.001	.029
	P ₈	.002	.046	.002	.043	.001	.037	.001	.025
	P ₉	.002	.046	.003	.054	.001	.031	.001	.033
	P ₁₀	.002	.045	.003	.054	.001	.033	.001	.032
Naïve malingering	P ₁₁	1	.000	1	.006	1	.000	1	.006
	P ₁₂	1	.000	1	1	1	.000	1	.006
	P ₁₃	1	.000	1	1	1	.000	1	.006
	P ₁₄	.007 ^a	.085 ^a	.008 ^a	.089 ^a	.005 ^a	.074 ^a	.003 ^a	.058 ^a
	P ₁₅	.008 ^a	.089 ^a	.010 ^a	.098 ^a	.005 ^a	.070 ^a	.003 ^a	.057 ^a
	P ₁₆	1	.000	1	1	1	.000	1	.000
	P ₁₇	1	.000	1	1	1	.000	1	.000
	P ₁₈	1	.000	1	1	1	.000	1	.000
	P ₁₉	1	.000	1	1	1	.000	1	.000
	P ₂₀	1	.000	1	1	1	.000	1	.000
	P ₂₁	1	.000	1	.017	.999	.029	1	.019
	P ₂₂	1	.000	1	1	1	.000	1	.000

Note: Samples from the posterior distribution = 27,000.

^aParticipants did not follow the role instructions.

The robustness analysis showed that this perfect performance was still achieved using three different priors on the malingering base rate. These priors on malingering base rates represent specific values and not probability distributions as in the case of Beta(5,5). These fixed values were selected considering prior knowledge about malingering base rates, based on the available literature. The prior on the malingering base rate of $\varphi = .3$ (i.e., 30%) was proposed following the Mittenberg and colleagues (2002) study with patients alleging personal injury, disability, and chronic pain. Priors on malingering base rates equivalent to

$\varphi = .4$ and $\varphi = .5$ were chosen according to Larrabee's reviews (2003, 2007). The base rate of $\varphi = .4$ (i.e., 40%) for mild traumatic brain injured litigants is commonly accepted among malingering researchers. The base rate of $\varphi = .5$ (i.e., 50%) might be appropriate for a wider range of civil and criminal settings.

Experiment 2: Clinical Group (Stroke) Versus Coached Malingerers

Method

The second experiment modified the first in two ways. One modification was to include a clinical sample of patients with demonstrable diagnosis of previous or acute stroke instead of using a bona fide group. This change was intended to give more external validity to our experiment and increase the possibility of extrapolating our findings to more realistic settings.

Another modification was to give special coaching as to how to complete the testing to all participants assigned to the malingering group. This group was called "coached malingering." The aim of this modification was to observe how coaching may affect the classification accuracy of both the below-chance criterion of SVT and the Bayesian approach. Evidence suggests that many people who undergo medical-legal assessments obtain some information about testing procedures, symptom presentation, or some strategies to avoid detection on psychological tests (Nicholson & Martelli, 2007). Except for these two modifications, the general procedure was identical to that from Experiment 1.

Participants. The clinical group was composed of 12 patients of the Stroke Unit of the Bethel Hospital of Bielefeld (Evangelisches Krankenhaus Bethel); men: $N = 8$; age $M = 68.50$ years; $SD = 14.93$, min = 35, max = 84 years; women: $N = 4$; age $M = 62.75$ years; $SD = 13.89$, min = 42, max = 71 years. Some participants had history of previous stroke; nevertheless, all of them were evaluated at the stroke unit days after a new incident or after presenting of suffering a new incident. Three participants suffered more than a single-stroke episode during the last 2 years. Stroke severity ranged from mild to severe. Etiology and location were also heterogeneous (e.g., ischemic posterior infarct right hemisphere [$N = 1$], lacunar infarct in the anterior limb of the internal capsule right hemisphere [$N = 1$] and left hemisphere [$N = 1$], middle cerebral artery stroke right hemisphere [$N = 3$] and left hemisphere [$N = 1$], thalamic infarct [$N = 1$], combined posterior and thalamic infarct right hemisphere [$N = 1$], and brainstem infarct right hemisphere [$N = 1$] and left hemisphere [$N = 1$]). The most common symptoms presented by the patients included sensorial (e.g., partial and complete hemianopsia, hypesthesia) and motor deficits (e.g., hemiparesis, dysarthria). The study was accomplished in compliance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

The "coached malingering" group was composed of 12 undergraduate students at Bielefeld University ($N = 12$; age: $M = 23.58$ years; $SD = 5.47$). Written informed consent was obtained from all participants previous to the experiment.

Procedure. Participants belonging to the clinical group were asked to give their best performance during testing. All patients were bedside evaluated using a 15" laptop on which the 45-trial forced-choice task of visual recognition was implemented. In contrast, coached malingering participants were evaluated in a standardized laboratory that was used in Experiment 1 as well. This group received explicit recommendations about how to face the tasks and feign cognitive impairments before evaluation. The information given to the coached malingering group was mainly: (a) do not exaggerate your cognitive impairment in a very obvious or implausible way; (b) some people with real cognitive impairment can also 'pass' this kind of task; (c) giving zero or just a few correct answers will be easy to detect; and (d) a random answering pattern will keep you at a chance level and make you more difficult to detect.

Measures. The same 45-trial computerized forced-choice task of visual recognition used in Experiment 1 was applied in the Coached Malingering condition. None of the participants of the coached malingering group had to be excluded due to neuropsychological deficits.

Results

Below-chance criterion of SVT. Total scores lower than 17 correct answers ($p < .05$) were considered as an indicator of poor effort during the task. Table 3 shows the obtained results for the coached malingering group. None of the 12 participants fell significantly below the chance level ($z = -2.09$). This result is in line with the hypothesis that suggests that coaching may diminish test sensitivity when the below-chance criterion is used to determine poor effort, as shown in Table 3. Nevertheless, again none of the participants of the bona fide group was misclassified. This again confirmed the high specificity of the below-chance criterion of SVT.

Table 3. Individual results of Experiment 2 within the clinical (stroke) and coached malingering groups using the below-chance criterion of SVT

Method	Clinical group (stroke)			Coached malingering group		
	Participant _i	Raw score	z-value	Participant _i	Raw score	z-value
SVT's below-chance criterion	P ₁	42	5.821	P ₁₃	36	3.876
	P ₂	41	5.522	P ₁₄	30	2.087
	P ₃	42	5.821	P ₁₅	27	1.193
	P ₄	45	6.716	P ₁₆	29	1.789
	P ₅	44	6.418	P ₁₇	24	0.596
	P ₆	43	6.119	P ₁₈	23	0.298
	P ₇	45	6.716	P ₁₉	26	1.193
	P ₈	44	6.418	P ₂₀	37	4.174
	P ₉	45	6.716	P ₂₁	24	0.596
	P ₁₀	44	6.418	P ₂₂	26	1.193
	P ₁₁	45	6.716	P ₂₃	26	1.193
	P ₁₂	45	6.716	P ₂₄	35	3.578

Note: SVT = symptom validity testing.

Table 4. Individual classification probabilities for Experiment 2 within the clinical (stroke) and coached malingering groups assuming different priors on malingering base rates

Posterior individual classification probabilities		Prior on malingering base rates							
		Beta(5,5)		$p(\varphi) = .5$		$p(\varphi) = .4$		$p(\varphi) = .3$	
Group	Participant _i	$p(z_i D)$		$p(z_i D)$		$p(z_i D)$		$p(z_i D)$	
		Mean z_i	SD z_i	Mean z_i	SD z_i	Mean z_i	SD z_i	Mean z_i	SD z_i
Clinical (stroke)	P ₁	.036	.187	.024	.154	.012	.109	.005	.071
	P ₂	.104	.305	.087	.281	.051	.220	.026	.160
	P ₃	.039	.193	.025	.155	.012	.111	.007	.083
	P ₄	.004	.060	.001	.031	.000	.014	.000	.015
	P ₅	.007	.081	.003	.053	.001	.035	.000	.014
	P ₆	.015	.120	.007	.085	.003	.050	.002	.042
	P ₇	.004	.066	.001	.033	.000	.022	.000	.011
	P ₈	.007	.082	.003	.054	.001	.029	.000	.020
	P ₉	.003	.057	.001	.030	.000	.016	.000	.006
	P ₁₀	.005	.072	.003	.051	.001	.035	.000	.018
	P ₁₁	.003	.053	.002	.041	.000	.014	.000	.012
	P ₁₂	.004	.062	.001	.028	.000	.019	.000	.012
Coached malingering	P ₁₃	.984	.127	.982	.134	.971	.169	.950	.219
	P ₁₄	1	.000	1	.012	1	.000	1	.011
	P ₁₅	1	.000	1	.000	1	.000	1	.000
	P ₁₆	1	.006	1	.006	1	.006	1	.006
	P ₁₇	1	.000	1	.000	1	.000	1	.000
	P ₁₈	1	.000	1	.000	1	.000	1	.000
	P ₁₉	1	.000	1	.000	1	.000	1	.000
	P ₂₀	.938	.241	.937	.244	.907	.290	.853	.355
	P ₂₁	1	.000	1	.000	1	.000	1	.000
	P ₂₂	1	.000	1	.000	1	.006	1	.006
	P ₂₃	1	.000	1	.000	1	.000	1	.000
	P ₂₄	.994	.078	.995	.074	.990	.098	.984	.126

Note: Samples from the posterior distribution = 27,000.

Bayesian latent group analysis of malingering. As in Experiment 1, we started the analysis using a relatively uninformative prior on the malingering base rate of Beta(5,5). This prior led to a mean posterior value for the base rate of .504, with a 95% credible interval ranging from .33 to .67.

Despite the coaching received by participants of the malingering group, individual estimates allowed to distinguish them from patients with stroke history. Table 4 shows that all participants within the clinical group showed low probabilities of being classified as coached malingerers. The highest individual mean value of $p(z_i|D)$ within the clinical group was 10.4%,

which certainly does not allow us to infer that the participant could belong to the malingering group. Within the coached malingering group, the classification probabilities $p(z_i|D)$ were very high for all the participants, with 93.8% being the lowest individual mean value of $p(z_i|D)$. These results suggest that a Bayesian latent group analysis of malingering appears to be resistant to the possible effects of coaching.

The robustness analysis considered the same fixed priors on malingering base rates of Experiment 1 (i.e., $\varphi = .5$, $\varphi = .4$, and $\varphi = .3$) and was performed to observe if there were significant variations in the classification variable z_i due to variations in the prior on base rates. The results showed little variations in the mean posterior values for the classification variable z_i when using different priors in the base rate. Consequently, the risk of misclassification can be considered very low. These results evidence the consistency of the Bayesian approach even when priors are varied across a plausible range. This is an interesting finding and may suggest the inclusion of Bayesian analysis as a complement to the existing methods, especially when coaching is suspected.

Discussion

During the past few years, Bayesian models have provided a formal framework that allows researchers to explore different ways to understand human cognition and to evaluate theoretical models against data (Griffiths, Kemp, & Tenenbaum, 2008; Lee, 2010). In our study, we examined whether a Bayesian approach could also be useful for the detection of poor effort and malingering in neuropsychological assessment.

Our results support the efficacy of a Bayesian latent group analysis for the detection of poor effort in the context of malingering. In both experiments, individual probabilities, $p(z_i|D)$, lead to the correct classification of all participants. Thus, these results show excellent levels of sensitivity and specificity.

These results become more important in the case of Experiment 2. This experiment introduced a clinical sample to assess whether our findings can be extrapolated to more realistic settings. Toward this goal, we also included specific coaching for the malingering group. Coaching is an important subject matter in malingering research, as previous studies have suggested that coaching may decrease the sensitivity of a test for poor effort (Brennan et al., 2009; Coleman, Rapport, Millis, Ricker, & Farchione, 1998; Russeler, Brett, Klaue, Sailer, & Munte, 2008; Suhr & Gunstad, 2000). Nevertheless, in our studies, the Bayesian approach was overall resistant to possible coaching effects of participants. Our finding becomes especially important when considering that sensitivity levels of the SVT tend to diminish after coaching (Dunn, Shear, Howe, & Ris, 2003; Russeler et al., 2008; Suhr & Gunstad, 2000). This fact again emphasizes the relevance of the decision rules (e.g., cutoff scores) used to determine the presence of poor effort during testing. In this context, the use of empirically obtained cutoff scores has contributed to increase sensitivity levels when determining malingering. Nevertheless, as will be pointed out later, we believe that our method might also contribute to this aim.

In terms of accuracy and precision, the individual posterior mean values for the classification variable z_i showed low levels of variation. Our analyses show that even with different priors on base rates, indices of variability and precision remain relatively stable, highlighting the robustness of the method.

From a frequentist point of view, the outcomes of our modeling can be considered as a form of cluster analysis, latent class analysis, or multivariate mixture estimation. In fact, it is often possible to use a frequentist statistical approach and obtain the same conclusions as Bayesian approaches. One clear advantage of the proposed Bayesian approach over the above-mentioned frequentist alternatives is the use of probabilities to quantify the uncertainty in one's judgment or knowledge, and so offer coherent, complete, and rational inference.

Likewise, receiver operating characteristic (ROC) curve analysis is another method that has been commonly used to estimate the classification accuracy of a particular test, but this is not the aim of the proposed method. However, one possible interpretation of ROC analysis to estimate the "probability that a randomly selected patient with the disorder has a test result indicating greater suspicion than of a randomly selected patient without the disorder" (Millis, 2008, p. 30) by observing the value of the area under the curve (i.e., AUC value). In this respect, our Bayesian approach can do that but can also accurately quantify the degree of confidence with which each particular participant (i.e., patient) is classified. This might constitute an advantage in comparison with the interpretation of a single AUC value. Another potential advantage of the presented Bayesian approach is that works independently of the test or the decision rule used (e.g., cutoff score). Therefore, raw scores obtained using validated SVTs (e.g., TOMM, WMT) can be used as input to accurately quantify our certainty on each person's classification.

As we showed in our two experiments, Bayesian methods might also constitute an interesting alternative for the assessment of effort in the field of applied neuropsychology. Recent studies suggest that it may be efficient to use multiple effort indicators (Boone, 2007, 2011; Larrabee, 2008, 2010; Larrabee et al., 2007) and multiple detection strategies (Neudecker & Skeel, 2009) when determining the presence of malingering. In line with this suggestion, we do not encourage our Bayesian latent group

approach as the one and only method to determine poor effort; instead, we believe that the Bayesian approach complements the information provided by the existing methods in useful ways. This synergy of methods may enrich the evaluation process and yield more reliable diagnoses of malingering.

Despite the overall success and robustness of our Bayesian approach, our analyses should be interpreted with some caution. First, the findings we presented are preliminary, and the use of a Bayesian latent group analysis for effort assessment is in an early stage. Second, in our experiments, the bona fide and stroke participants behaved very differently from both groups of participants who were instructed to malingering (with or without coaching); therefore, the malingerers were relatively easy to identify. When participants are difficult to classify the z_i variable should fall in the intermediate range. Third, even though it is well documented that Bayesian methods can be applied in small samples, it must be noted that classification accuracy indices (i.e., sensitivity and specificity) will most likely shrink upon cross-validation when using larger samples. This can be considered as a limitation of our study. Additionally, it needs to be considered that variations in the base rates (see Streiner, 2003, p. 214) and the selected gold standard (Millis, 2008) may also affect classification accuracy indices.

Future research could replicate our findings perhaps using base rates on malingering that are lower than the ones considered here. For instance, it would be interesting to consider the base rates found in some large consecutive clinical samples that include bona fide patients as well as participants with feigned or exaggerated symptoms (Donders & Strong, 2011).

In conclusion, our data and analyses showed the potential utility of a Bayesian approach for effort assessment and the detection of malingering (see also Iverson, 2007). It is unfortunate that, despite their prominence in fields such as statistics and machine learning, Bayesian methods are still rarely used in mainstream clinical practice. Our Bayesian model is simple to use and yields intuitive conclusions. First and foremost, it attaches to each individual a probability of malingering. We hope that Bayesian methods such as ours can be used in combination with existing techniques to draw conclusions about poor effort and malingering that are both intuitive and informative.

Supplementary material

Supplementary material is available at *Archives of Clinical Neuropsychology* online.

Funding

This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft) (EC 277 to HJM and MP) and the Dutch Organization for Scientific Research (NOW) (Vidi grant to E-JW).

Conflict of Interest

None declared.

Acknowledgements

We thank our colleagues in the Physiological Psychology Department and the CITEC Research Groups at Bielefeld University for their expert scientific and technical support and helpful advice. Moreover, we especially wish to thank Prof. Dr med. Wolf-Rüdiger Schäbitz and the team of the Stroke Unit in the Clinic for Neurology of the Bethel Hospital of Bielefeld (Evangelisches Krankenhaus Bethel) for their help and assistance when carrying out our clinical study.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: Author.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (Revised 4th ed.). Washington, DC: Author.
- Beetar, J. T., & Williams, J. M. (1995). Malingering response styles on the memory assessment scales and symptom validity tests. *Archives of Clinical Neuropsychology*, 10, 57–72.
- Binder, L. M. (1993). *Portland Digit Recognition Test manual* (2nd ed.). Portland, OR: Private Publication.
- Boone, K. B. (2007). A reconsideration of the Slick et al. criteria for malingered neurocognitive dysfunction. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 29–49). New York: Guilford Press.
- Boone, K. B. (2011). *Assessment of response bias: Embedded effort indicators*. Presentation at the Second European Symposium on Symptom Validity Assessment, London.
- Brennan, A., Meyer, S., David, E., Pella, R., Hill, B., & Gouvier, W. D. (2009). The vulnerability to coaching across measures of effort. *The Clinical Neuropsychologist*, 23, 314–328.

- Bush, S. S., Ruff, R. M., Troster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., et al. (2005). Symptom validity assessment: Practice issues and medical necessity—NAN policy & planning committee. *Archives of Clinical Neuropsychology*, *20*, 419–426.
- Clemens, I. A. H., De Vrijer, M., Selen, L. P. J., Van Gisbergen, J. A. M., & Medendorp, W. P. (2011). Multisensory processing in spatial orientation: An inverse probabilistic approach. *Journal of Neuroscience*, *31*, 5365–5377.
- Coleman, R. D., Rapport, L. J., Millis, S. R., Ricker, J. H., & Farchione, T. J. (1998). Effects of coaching on detection of malingering on the California Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology*, *20*, 201–210.
- Denney, R. L. (2008). Negative response bias and malingering during neuropsychological assessment in criminal forensic settings. In R. L. Denney, & J. P. Sullivan (Eds.), *Clinical neuropsychology in the criminal forensic setting* (pp. 91–134). New York: Guilford Press.
- Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, *6*, 274.
- Donders, D. J., & Strong, C. A. H. (2011). Embedded effort indicators on the California Verbal Learning Test - Second Edition (CVLT II): An attempted cross-validation. *The Clinical Neuropsychologist*, *25*, 173–184.
- Dunn, T. M., Shear, P. K., Howe, S., & Ris, M. D. (2003). Detecting neuropsychological malingering: Effects of coaching and information. *Archives of Clinical Neuropsychology*, *18* (2), 121–134.
- Frederick, R.I. (1997). *Validity Indicator Profile manual*. Minnetonka, MN: NCS Assessments.
- Frederick, R. I., & Speed, F. M. (2007). On the interpretation of below-chance responding in forced-choice tests. *Assessment*, *14*, 3–11.
- Gamerman, D., & Lopes, H. F. (2006). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*, 2nd Ed. (Texts in Statistical Science Series). New York: Chapman & Hall/CRC.
- Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*. Interdisciplinary Statistics, London: Chapman & Hall.
- Green, P. (2003). *Word Memory Test for Windows: Test manual*. Edmonton, Alberta: Green's Publishing.
- Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *Cambridge handbook of computational Psychology* (pp. 59–100). Cambridge: Cambridge University Press.
- Grote, C. L., & Hook, J. N. (2007). Forced-Choice recognition tests of malingering. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 44–79). Oxford, NY: Oxford University Press.
- Horton, A. M., Jr. (2010). Overview of forensic neuropsychology. In , Jr.A. Horton, & L. Hartlage (Eds.), *Handbook of forensic neuropsychology* (2nd ed., pp. 3–9). New York: Springer Publishing.
- Iverson, G. (2003). Detecting malingering in civil forensic evaluations. In A. H. Mac-Neill, & L. C. Hartlage (Eds.), *Handbook of forensic neuropsychology* (1st ed., pp. 137–177). New York: Springer Publishing.
- Iverson, G. (2007). Identifying exaggeration and malingering. *Pain Practice*, *7*, 94–102.
- Iverson, G., & Binder, L. (2000). Detecting exaggeration and malingering in neuropsychological assessment. *Journal of Head Trauma Rehabilitation*, *15*, 829–858.
- Jarvis, W. (2008). *Medialab Version 2008.1. 22 and DirectRT Version 2008.1. 0.11. Computer software*. New York: Empirisoft.
- Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge: Cambridge University Press.
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Clarendon Press.
- Kruschke, J. K., & John, K. (2010). *Doing Bayesian data analysis: A tutorial introduction with R and BUGS*. Burlington, MA: Academic Press/Elsevier.
- Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, *17*, 410–425.
- Larrabee, G. J. (2007). Malingering, research resists, and base rates. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (xiv+386 pp.). Oxford, NY: Oxford University Press.
- Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist*, *22*, 666–679.
- Larrabee, G. J. (2010). *Performance validity and test validity in neuropsychological assessment*. Presentation at the workshop of the International Academy of Applied Neuropsychology, London.
- Larrabee, G. J., Greiffenstein, M., Grewe, K., & Bianchini, K. (2007). Refining diagnostic criteria for malingering. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 334–372). Oxford, NY: Oxford University Press.
- Larsen, J. D., Allen, M. D., Bigler, E. D., Goodrich-Hunsaker, N. J., & Hopkins, R. O. (2010). Different patterns of cerebral activation in genuine and malingered cognitive effort during performance on the Word Memory Test. *Brain Injury*, *24*, 89–99.
- Lee, M. D. (2010). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology*, *55*, 1–7.
- Lee, M. D., & Wagenmakers, E. J. (2010). *A course in Bayesian graphical modeling for cognitive science: Unpublished course materials*. Retrieved October 05, 2010, from <http://www.ejwagenmakers.com/BayesCourse/BayesBookWeb.pdf>.
- Lee, T. M. C., Au, R. K. C., Liu, H. L., Ting, K. H., Huang, C. M., & Chan, C. C. H. (2009). Are errors differentiable from deceptive responses when feigning memory impairment? An fMRI study. *Brain and Cognition*, *69*, 406–412.
- Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*, *28*, 3049–3067.
- Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325–337.
- Macmillan, N., & Creelman, C. (2008). *Detection theory: A user's guide* (2nd ed.). New Jersey: Lawrence Erlbaum.
- Miller, L., Boyd, M., Cohn, A., Wilson, J., & McFarland, M. (2006). *Prevalence of sub-optimal effort in disability applicants*. Paper presented at the 34th annual meeting of the International Neuropsychological Society, Boston.
- Millis, S. (2008). What clinicians really need to know about symptom exaggeration, insufficient effort, and malingering: Statistical and measurement matters. In J. E. Morgan, & J. J. Sweet (Eds.), *Neuropsychology of malingering casebook* (pp. 21–37). Hove, East Sussex: Psychology Press.
- Mittenberg, W., Patton, C., Canyock, E., & Condit, D. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, *24*, 1094–1102.
- Morel, K. R., & Shepherd, B. E. (2008). Developing a symptom validity test for posttraumatic stress disorder: Application of the binomial distribution. *Journal of Anxiety Disorders*, *22*, 1297–1302.

- Mossman, D. (2000). The meaning of malingering data: Further applications of Bayes' theorem. *Behavioral Sciences and the Law*, 18, 761–779.
- Mossman, D., & Hart, K. J. (1996). Presenting evidence of malingering to courts: Insights from decision theory. *Behavioral Sciences and the Law*, 14, 271–291.
- Neudecker, J. J., & Skeel, R. L. (2009). Development of a novel malingering detection method involving multiple detection strategies. *Archives of Clinical Neuropsychology*, 24, 59–70.
- Nicholson, K., & Martelli, M. F. (2007). Malingering: Overview and basic concepts. In G. Young, A. W. Kane, K. Nicholson, & D. W. Shuman (Eds.), *Causality of psychological injury: Presenting evidence in court* (xi+646 pp.). New York: Springer.
- Nilsson, H., Rieskamp, J., & Wagenmakers, E.-J. (2010). Hierarchical Bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology*, 55, 84–93.
- O'Hagan, A., & Forster, J. J. (2004). *Kendall's Advanced Theory of Statistics, Volume 2B: Bayesian inference* (2nd ed.). London: Arnold.
- Pankratz, L., Fausti, S. A., & Peed, S. (1975). Forced-choice technique to evaluate deafness in hysterical or malingering patient. *Journal of Consulting and Clinical Psychology*, 43, 421–422.
- Piefke, M., Onur, A., & Fink, G. R. (2010). Aging-related changes of neural mechanisms underlying visual-spatial working memory. *Neurobiology of Aging*. PMID: 21130531.
- Piefke, M., Weiss, P.H., Zilles, K., Markowitsch, H.J., & Fink, G.R. (2003). Differential remoteness and emotional tone modulate the neural correlates of autobiographical memory. *Brain*, 126, 650–668.
- Poirier, D. J. (2006). The growth of Bayesian methods in statistics and economics since 1970. *Bayesian Analysis*, 1, 969–980.
- Rogers, R. (2008). Detection strategies for malingering and defensiveness. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., pp. 14–35). New York: Guilford Press.
- Rogers, R., & Salekin, R. T. (1998). Research report beguiled by Bayes: A re-analysis of Mossman and Hart's estimates of malingering. *Behavioral Sciences and the Law*, 16, 147–153.
- Rogers, R., Salekin, R. T., Sewell, K., Goldstein, A., & Leonard, K. (1998). A comparison of forensic and nonforensic malingerers: A prototypical analysis of explanatory models. *Law and Human Behavior*, 22, 353–367.
- Rogers, R., Sewell, K. W., & Goldstein, A. M. (1994). Explanatory models of malingering: A prototypical analysis. *Law and Human Behavior*, 18, 543–552.
- Russeler, J., Brett, A., Klaue, U., Sailer, M., & Munte, T. F. (2008). The effect of coaching on the simulated malingering of memory impairment. *BMC Neurology*, 8, 1–14.
- Samaniego, F. J. (2010). *A comparison of the Bayesian and Frequentist approaches to estimation*. New York: Springer.
- Schulte-Rüther, M., Markowitsch, H. J., Fink, G. R., & Piefke, M. (2007). Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: A functional magnetic resonance imaging approach to empathy. *Journal of Cognitive Neuroscience*, 19, 1354–1372.
- Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E. J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, 32, 1248–1284.
- Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, 13, 545–561.
- Slick, D. J., Tan, J. E., Strauss, E., Mateer, C. A., Harnadek, M., & Sherman, E. M. (2003). Victoria Symptom Validity Test scores of patients with profound memory impairment: Nonlitigants case studies. *The Clinical Neuropsychologist*, 17, 390–394.
- Stephens, M., & Balding, D. J. (2009). Bayesian statistical methods for genetic association studies. *Nature Reviews Genetics*, 10, 681–690.
- Streiner, D. (2003). Diagnosing tests: Using and misusing diagnostic and screening tests. *Journal of Personality Assessment*, 81, 209–219.
- Suhr, J. A., & Gunstad, J. (2000). The effects of coaching on the sensitivity and specificity of malingering measures. *Archives of Clinical Neuropsychology*, 15, 415–424.
- Tombaugh, T. N. (1996). *Test of Memory Malingering TOMM*. New York: Multi-Health Systems.
- Wolfe, P. L., Millis, S. R., Hanks, R., Fichtenberg, N., Larrabee, G. J., & Sweet, J. J. (2010). Effort indicators within the California verbal learning test-II (CVLT-II). *The Clinical Neuropsychologist*, 24, 153–168.