

Cognitive model decomposition of the BART: Assessment and application

Don van Ravenzwaaij*, Gilles Dutilh, Eric-Jan Wagenmakers

University of Amsterdam, The Netherlands

ARTICLE INFO

Article history:

Received 24 November 2009

Received in revised form

21 June 2010

Available online 17 September 2010

Keywords:

BART

Hierarchical Bayesian modeling

Cognitive model

Risk taking

Alcohol

ABSTRACT

The Balloon Analogue Risk Task, or BART, aims to measure risk taking behavior in a controlled setting. In order to quantify the processes that underlie performance on the BART, Wallsten, Pleskac, and Lejuez (2005) proposed a series of mathematical models whose parameters have a clear psychological interpretation. Here we examine a 2-parameter simplification of Wallsten et al.'s preferred 4-parameter model. A parameter recovery study shows that – with plausible restrictions on the number of participants and trials – both parameters (i.e., risk taking γ^+ and response consistency β) can be estimated accurately. To demonstrate how the 2-parameter model can be used in practice, we implemented a Bayesian hierarchical version and applied it to an empirical data set in which participants performed the BART following various amounts of alcohol intake.

© 2010 Elsevier Inc. All rights reserved.

When people take a risk, they pursue some form of reward while exposing themselves to potential harm (Wallach, Kogan, & Bem, 1962). Depending on the situation, such harm can include bankruptcy, cocaine addiction, sexually transmitted diseases, and even death. Among the factors affecting risk taking behavior – why some people take risks when other people decide to play it safe – a very influential one is substance abuse (Adlaf & Smart, 1983), of which a common example is the abuse of alcohol.

The effects of alcohol on risk taking have been the topic of extensive research. Among others, alcohol abuse has been found to increase risk taking during driving (e.g., Burian, Liguori, & Robinson, 2002; Cohen, Dearnaley, & Hansel, 1958), reduce the perceived negative consequences of risk taking (e.g., Fromme, Katz, & D'Amico, 1997), increase the participation in unsafe sex (e.g., McEwan, McCallum, Bhopal, and Madhok (1992) and Kalichman, Heckman, and Kelly (1996), but see Leigh and Stall (1993) for cautionary remarks) and increase the number of accidents encountered (e.g., Cherpitel, 1993a,b). In this paper, we experimentally investigate the effects of three doses of alcohol on risk taking behavior.

The study of risk taking generally proceeds along one of two research traditions. The first tradition is the most direct in that it uses self-report questionnaires to measure risk-related tendencies such as impulsivity and sensation seeking (e.g., Eysenck & Eysenck, 1977). While the direct approach provides a measurement of risk

taking that is straightforward and transparent, the fact that this approach relies on self-report means that the results hinge on the truthfulness of the respondent. Because respondents might not answer accurately for a variety of reasons (e.g., shame, insufficient self-knowledge, fear of consequences, e.g., Edwards (1957)), it is desirable to have other ways to measure risk taking.

The second research tradition is less direct, as it uses experimental tasks to measure risk taking behavior in a controlled setting. One such experimental task is the Balloon Analogue Risk Task, or BART (Lejuez et al., 2002). On every trial of the BART, a computer screen displays a balloon that represents a small monetary value (see Fig. 1). The participant is presented with a choice; the first option is to play it safe and secure the amount of money the balloon is worth by transferring the money to a virtual bank account (i.e., *cash*). The second option is to take a risk and add a small amount of air to the balloon (i.e., *pump*).

When the participant pumps, the balloon can burst, and all the money that the balloon represents is lost. However, when the balloon is pumped and does not burst, it grows in size and is worth more money—when this happens, the participant is again confronted with the choice: cash or pump. A new trial begins when the participant cashes or the balloon bursts.

In the original version of the BART (Lejuez et al., 2002), the probability that the balloon bursts, p^{burst} , increases with every pump according to

$$p^{\text{burst}} = \frac{1}{x - n^{\text{pumps}}} \quad \text{with } x > 0, \quad (1)$$

where n^{pumps} is the number of pumps in the trial so far and x is a positive integer determined by the experimenter. In their original paper, Lejuez et al. (2002) used $x = 8$, $x = 32$, and $x = 128$.

* Corresponding address: University of Amsterdam, Department of Psychology, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands.

E-mail address: d.vanravenzwaaij@uva.nl (D. van Ravenzwaaij).

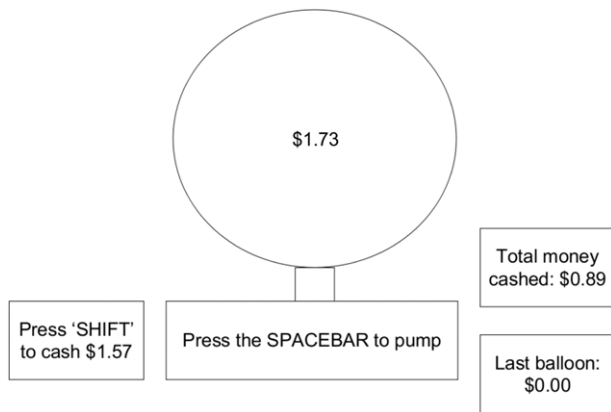


Fig. 1. A screenshot of the Balloon Analogue Risk Task (BART).

The BART is a simple laboratory task that nonetheless captures the defining characteristic of risk taking in the real world—when participants pump the balloon, they pursue reward while exposing themselves to potential harm. For the remainder of this paper, we will be using a version of the BART in which p^{burst} is fixed over pump opportunities and the average gain of every pump decision is exactly 0. This will allow us to look at the effects of risk-taking in isolation.

Performance on the BART is usually quantified by the mean number of pumps across trials, excluding balloons that burst. This measure has been shown to correlate with self-reported risk taking behaviors such as alcohol abuse, smoking, drug abuse, gambling, unsafe sex, and even stealing (e.g., Hopko et al., 2005; Lejuez, Aklin, Zvolensky, & Pedulla, 2003; Lejuez et al., 2002). Little is known, however, about the cognitive processes that cause suboptimal performance on the BART. For instance, drug addicts might perform poorly on the BART because they take too much risk; alternatively, they may have trouble learning from experience, or they may be more erratic when it comes to translating their preference into action. Based on the observed data alone, these different possibilities cannot be distinguished.

One way to learn more about the unobserved psychological processes that determine performance of the BART is with the use of a cognitive process model. Cognitive process models propose concrete cognitive mechanisms that underlie observed behavior; therefore, a cognitive process model is a means to translate what is observed but relatively uninformative to what is unobserved and relatively informative. An example of a successful cognitive process model in the study on risk taking is the Expectancy–Valence model for the Iowa gambling task (e.g., Busemeyer & Stout, 2002; Wetzels, Vandekerckhove, Tuerlinckx, & Wagenmakers, 2010).

In an attempt to increase understanding of the psychological processes involved in the BART, Wallsten et al. (2005) proposed a series of cognitive process models. These models include parameters that quantify risk taking (the psychological process of interest), speed of learning from experience, and behavioral consistency. With the help of these models, researchers can study the risk taking process separately from other psychological processes that together determine performance on the BART. Thus, the BART models proposed by Wallsten et al. allow a decomposition of observed behavior into its constituent cognitive processes. Unfortunately, the BART models have not been applied often, and consequently not much is known about how well the models are able to estimate the processes they purport to measure.

The goals of this paper are twofold. First, this paper seeks to increase knowledge about how the BART models can be applied to data. In order to do so, we will assess parameter recovery of a simplified version of the BART model not originally considered

by Wallsten et al. (2005). Second, this paper seeks to investigate the effects of alcohol intake on the BART. We will analyse the experimental data with a Bayesian hierarchical implementation of the BART model.

The remainder of this paper is organized in five sections. In the first section we introduce the BART models. In the second section we discuss Bayesian modeling and the extension to hierarchical Bayesian modeling. In the third section we present a number of simulations that seek to establish whether the model can recover the parameter values that were used to generate simulated data. In the fourth section we present experimental data and fit a Bayesian hierarchical implementation of a 2-parameter simplification of the model to a data set in which we manipulated alcohol intake prior to administration of the BART, after which the last section concludes.

1. The BART models

The BART models by Wallsten et al. (2005) are cognitive decision models inspired by the Expectancy–Valence model (Busemeyer & Stout, 2002) for the famous Iowa Gambling Task (Bechara, Damasio, Tranel, & Damasio, 1997). In their article, Wallsten et al. presented a total of 10 models that make different assumptions about the details of the decision process (for an overview see Wallsten et al. (2005), p. 870, Table 2). As a basis of our discussion we use Wallsten et al.'s “Model 3”, a parsimonious model that fit the data relatively well (see Wallsten et al. (2005), p. 872, Table 3). This parsimonious model has 4 parameters and will from here on be called the 4-parameter model. Apart from the 4-parameter model, we assessed parameter recovery of two 3-parameter simplifications and one 2-parameter simplification that were all not originally considered by Wallsten et al. (2005).

The 4-parameter model assumes that, on a particular trial k , the decision maker (henceforth DM) believes that there is a single, constant probability that a pump will make the balloon burst, p_k^{belief} . Thus, on any given trial DM is assumed to believe that the balloon is just as likely to burst after the first pump than after, say, the fifth pump. According to this model, DM starts the first trial with a prior belief about the probability that a pump will make the balloon burst, a prior belief that is updated on subsequent trials:

$$p_k^{belief} = 1 - \frac{\alpha + \sum_{K=0}^{k-1} n_K^{success}}{\mu + \sum_{K=0}^{k-1} n_K^{pumps}} \quad \text{with } \alpha < \mu. \quad (2)$$

In this updating equation, the quantity $1 - \alpha/\mu$ reflect DM's prior belief that pumping the balloon will make it burst. The absolute size of α and μ determines the rate with which DM learns from the data, with higher values indicating that more data is needed to overwhelm DM's prior belief. The quantity $\sum_{K=0}^{k-1} n_K^{success}$ is the number of successful (non-bursting) pumps up to trial k , and $\sum_{K=0}^{k-1} n_K^{pumps}$ is the total number of pumps up to trial k .

As an example, consider participant Jack. At the beginning of the experiment, Jack's α is 18 and Jack's μ is 20. This means his prior belief that pumping the balloon will make it burst equals $1 - 18/20 = 0.1$. Suppose that on the first trial, Jack pumps twice, and then the balloon bursts. This means Jack's new belief about the bursting probability is $1 - (18 + 1)/(20 + 2) = 0.136$. Consequently, Jack will pump more cautiously in the future (see also Eqs. (3) and (4) below).

The next assumption is that DM determines the number of pumps prior to the first pump, and does not make adjustments during pumping. The number of pumps that DM considers optimal on trial k , ω_k , depends both on DM's propensity for risk taking, γ^+ ,

and DM's belief about the probability that pumping the balloon will make it burst:

$$\omega_k = \frac{-\gamma^+}{\ln(1 - p_k^{\text{belief}})} \quad \text{with } \gamma^+ \geq 0. \quad (3)$$

The actual probability that DM will pump on trial k for pump opportunity l , p_{kl}^{pump} , depends both on the number of pumps DM considers optimal, ω_k , and on DM's behavioral consistency β :

$$p_{kl}^{\text{pump}} = \frac{1}{1 + e^{\beta(l - \omega_k)}} \quad \text{with } \beta \geq 0. \quad (4)$$

This logistic equation shows that high values for β mean less variable responding. When $\beta = 0$, $p_{kl}^{\text{pump}} = 0.5$, and DM's decision to pump or to cash is random. When $\beta \rightarrow \infty$, DM's behavior is completely determined by whether or not the pump opportunity l exceeds the number of pumps that DM considers optimal: if $l - \omega_k > 0$ (i.e., the optimum number has been exceeded), $p_{kl}^{\text{pump}} \rightarrow 0$, and DM is virtually certain to stop pumping; if $l - \omega_k < 0$ (i.e., the optimum number has not yet been reached), $p_{kl}^{\text{pump}} \rightarrow 1$, and DM is virtually certain to continue pumping.

In order to fit the 4-parameter BART model to observed data and infer the parameter values that are most consistent with DM's performance, the parameters are connected to the data via the likelihood function. For the 4-parameter model, the probability of the data, $p(D|\alpha, \mu, \gamma^+, \beta)$ for all trials, n_k , and for all pumps within each trial, $n_{l(k)}$, depends on the probability that DM will pump for each trial k for each pump opportunity l

$$p(D|\alpha, \mu, \gamma^+, \beta) = \prod_{k=1}^{n_k} \prod_{l=1}^{n_{l(k)}} p_{kl}^{\text{pump}} (1 - p_{k, n_{l(k)}+1}^{\text{pump}})^{d_k}, \quad (5)$$

where $d_k = 1$ if DM cashed on trial k and $d_k = 0$ if the balloon burst on trial k . This quantity is basically the product of all probabilities that DM will pump times one minus this probability on the occasions where DM cashed. The likelihood of the parameters given the data is proportional to the probability of the data given the parameters (e.g., Edwards, 1992; Myung, 2003), so that $L(\alpha, \mu, \gamma^+, \beta|D) \propto p(D|\alpha, \mu, \gamma^+, \beta)$. The parameters to be estimated for the 4-update model are α, μ, γ^+ , and β .

So far, we have dealt exclusively with the 4-parameter model. However, our simulations – reported below and in an online Appendix¹ – will indicate that this model needs to be simplified. In the online Appendix we considered the 4-parameter model and two 3-parameter simplifications, here we only discuss one 2-parameter simplification, because it was the model that most accurately recovered its parameters.

The 2-parameter model assumes that DM's belief about the probability that pumping the balloon will make it burst is fixed over trials. In other words, DM does not learn. This means we can drop the subscript k from p_k^{belief} ; parameter p^{belief} is now fixed and does not need to be estimated. This is a realistic model when the participant is told the actual bursting probability in advance. The parameters to be estimated for the 2-parameter model are γ^+ and β .

2. Bayesian parameter estimation

In previous work, parameter estimation for the 4-parameter BART model was carried out by means of individual subject maximum likelihood (Wallsten et al., 2005).² This means that the model was applied to each participant's data separately, and that inference concerned the parameter point values that make the observed data most likely.

Here we estimate the parameters of the BART model in a Bayesian way. In Bayesian inference, the researcher starts with prior probability distributions, or *priors*, that reflect the researcher's uncertainty or belief about the parameters before the data have been observed. In the next step, the prior distributions are updated by means of the data (i.e., the likelihood), and the result is a joint *posterior* distribution for the model parameters. This posterior distribution reflects the researcher's uncertainty or degree of belief about the parameters after the data have been observed.

Bayesian inference has several advantages over maximum likelihood (e.g., Carlin & Louis, 2000; Congdon, 2010; Wagenmakers, Lee, Lodewyckx, & Iverson, 2008). First, modern Bayesian parameter estimation techniques make it easy to extend a Bayesian model to handle realistic situations in which structure is added at the group level (e.g., random effects, mixtures, and contaminants). Second, Bayesian model selection procedures allow researchers to quantify the support that the data provide both for and against a null hypothesis (Carlin & Louis, 2000; Gallistel, 2009; Rouder, Speckman, Sun, Morey, & Iverson, 2009; Wetzels, Raaijmakers, Jakab, & Wagenmakers, 2009); in a manner similar to parameter estimation, Bayesian model selection starts by specifying prior probabilities for each of the competing models. The prior probabilities are updated through the data to yield posterior probabilities for the competing models. Third, posterior distributions automatically and naturally give an idea about the uncertainty in the inference (Congdon, 2010).

2.1. Hierarchical extension

As mentioned above, Bayesian models can be easily extended to more realistic scenarios, such as those that feature group-level structure, and, in particular, random effects that describe individual differences. Historically, the field of experimental psychology has mostly ignored individual differences, tacitly assuming that each new participant is a replicate of the previous one (Batchelder, 2007). As Estes and others have shown, however, individual differences that are ignored can lead to averaging artifacts, where the inference for the grouped data is no longer representative for any of the participants (e.g., Estes, 1956, 2002; Heathcote, Brown, & Mewhort, 2000). One way to address this issue, popular in psychophysics, is to measure each individual participant extensively, and analyze the data on a participant-by-participant basis.

In between the two extremes of assuming that participants are completely the same and that they are completely different lies the compromise of hierarchical modeling (see also Lee & Webb, 2005; Nilsson, Rieskamp, & Wagenmakers, 2011). In hierarchical modeling, individual parameters are assumed to be drawn from an overarching group distribution (Gelman & Hill, 2007). This group distribution has parameters of its own, called *hyperparameters*. Usually, one starts by assuming that individual-level parameters are constrained by a Gaussian group distribution, $N(\mu, \sigma)$; because σ corresponds to the spread of the group distribution, this parameter quantifies the extent to which the participants differ—low values of σ indicate that the participants are relatively similar; in the limit of $\sigma \rightarrow 0$, all participants are identical copies of each other.

The theoretical advantages and practical relevance of a Bayesian hierarchical analysis for common experimental designs has been repeatedly demonstrated by Jeff Rouder and colleagues (e.g., Rouder & Lu, 2005; Rouder, Lu, Morey, Sun, & Speckman, 2008; Rouder, Lu, Speckman, Sun, & Jiang, 2005; Rouder et al., 2007). One of the theoretical advantages is that by hierarchical modeling, researchers automatically obtain an optimal compromise between the extremes of complete pooling and complete independence. This approach is used by a number of authors in this

¹ See <http://www.donvanravenzwaaij.com>, section "Research & Codes".

² Matlab code can be found on the web page of Tim Pleskac.

special issue, including Merkle, Smithson, and Verkuilen (2011), and Nilsson et al. (2011), who also consider individual differences in simple decision-making models. One of the practical advantages is that hierarchical modeling allows for more efficient inference on the individual level; this happens because extreme individual estimates, when these are based on few data, are shrunk towards the group mean (Gelman & Hill, 2007).

2.2. Implementation

We implemented the BART models in the WinBUGS environment (Lunn, Spiegelhalter, Thomas, & Best, 2009; Lunn, Thomas, Best, & Spiegelhalter, 2000; Ntzoufras, 2009), of which introductions for psychologists are given by Lee and Wagenmakers (2009) and Sheu and O'Curry (1998). WinBUGS is a general-purpose program that allows users to specify and fit a wide array of Bayesian models. Although WinBUGS does not work for every application, it will work for most applications in psychology. The WinBUGS program is easy to learn and is supported by a large community of active researchers. In WinBUGS, the user needs to specify the model (i.e., the likelihood and the priors—see the Appendix for an example), and provide the model with the data. Next, the WinBUGS program uses Markov chain Monte Carlo (MCMC) to draw values from the posterior distribution. One of the advantages of WinBUGS is that the user does not need to hand-code the MCMC algorithms, as these are applied by WinBUGS per default.

An example of the kind of inference that WinBUGS affords is shown in Fig. 2. The top panel of this figure shows three chains that are designed to sample values from the posterior for γ^+ , the risk taking parameter for a participant in the drunk condition of our experiment (explained in detail later). The figure shows that each chain samples 5000 values; a first set of 5000 values was discarded as burn-in, to eliminate any dependence on the starting values of the chains. Visual inspection shows that the three chains are virtually indistinguishable — this indicates that the chains are drawing samples from the same distribution — and that the chains do not exhibit slow upward or downward trends — this indicates that the distribution is sampled efficiently.

To confirm more formally that the three chains have converged to the posterior distribution, one method is to calculate the \hat{R} statistic (Gelman & Rubin, 1992), a statistic that compares the variance over chains to that within chains. When the chains are indistinguishable, \hat{R} equals 1. As a rule of thumb, an \hat{R} higher than 1.10 is considered suspicious. For the three chains shown in Fig. 2, $\hat{R} = 1.00$.

Having reassured ourselves that the three chains draw samples from the posterior distribution, we can then pool the samples and plot these as a histogram. The result is shown in the bottom panel of Fig. 2. Based on the 15,000 samples, we can also construct a Bayesian 95% confidence interval (also known as *credible interval*). In this case, after seeing the data, we can be 95% confident that γ^+ lies in the interval (0.78, 1.12). In addition, we can summarize the posterior distribution by its mean, median, or mode.

3. Parameter recovery simulations

In this section we examine parameter recovery of the 2-parameter simplification of the BART model. In this model, p^{belief} is fixed to the value of p^{burst} . This way, the only parameters left to estimate are γ^+ and β . We generated data for a grid of values for parameters γ^+ and β . We did so by plugging in the parameter values into Eqs. (3) and (4), to calculate the probability that DM will pump on trial k for pump opportunity l , p_{kl}^{pump} . We generated pumps and cashes based on these probabilities, and also incorporated bursting probability after each pumping decision to

Table 1

Parameter correlations in the 2-parameter model with 90 trials per simulation.

		β				
		0.4	0.5	0.6	0.7	0.8
γ^+	0.6	−0.79	−0.72	−0.75	−0.73	−0.73
	1	−0.82	−0.82	−0.82	−0.79	−0.71
	1.4	−0.85	−0.83	−0.81	−0.74	−0.69
	1.8	−0.85	−0.82	−0.77	−0.72	−0.67
	2.2	−0.82	−0.76	−0.73	−0.65	−0.54

generate a full dataset. Next, we fit the model to the simulated data and compared the resulting parameter estimates (specifically, the posterior mean) with the original values that were used to generate the data.

For all simulations, parameters were recovered with a Bayesian implementation of the model, WinBUGS code of which can be found in the Appendix. We used the following priors: $\gamma^+ \sim U(0, 10)$ and $\beta \sim U(0, 10)$, where U indicates the uniform distribution. In the absence of strong prior knowledge about the parameters, these priors were chosen to be relatively vague. Other vague priors (e.g., uniform priors on the log of γ^+ and β) yielded similar results. For each of the model fits in the next section, we used a single chain, consisting of 2000 iterations with a burn-in of 1000 samples. The simulations were conducted with a range of starting values for the MCMC chains. The results were qualitatively similar, unless reported otherwise. Parameter p^{burst} was set to 0.15, so that p^{belief} was also 0.15. We ran 1000 simulations of 1 participant completing 90 trials.

Fig. 3 shows recovery of the γ^+ parameter. Recovery of γ^+ is good for most combinations of true values for γ^+ and β . For a combination of low γ^+ and low β parameters, the estimate of γ^+ becomes too high. Also, the estimates become more variable.

Fig. 4 shows recovery of the β parameter. Recovery of β is good for the whole range of values for the γ^+ and β parameters, although there is a small bias for the extreme values of γ^+ .

Table 1 presents the correlation between the different parameters for the posterior means. Analogous to the more complicated models, this table shows that γ^+ and β estimates are highly negatively correlated. This correlation seems to be high when β takes a low value and when γ^+ takes an intermediate value. Regardless, there is a substantial parameter dependency.

In order to examine the effect of burst probability on parameter recovery, we simulated data for the following range of values for p^{burst} : 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45 and “variable”. The “variable” p^{burst} meant that for one third of the trials, $p^{\text{burst}} = 0.1$, for one third $p^{\text{burst}} = 0.15$, and for one third $p^{\text{burst}} = 0.2$. We included the “variable” p^{burst} in our simulations because it is identical to p^{burst} that was used in the experiment presented below. Since $p^{\text{belief}} = p^{\text{burst}}$, we simultaneously varied p^{belief} . Analogous to the other simulations, $\gamma^+ = 1.4$ and $\beta = 0.6$. We ran 1000 simulations of 1 participant completing 90 trials. The results of this simulation are displayed in Fig. 5.

Recovery of the γ^+ parameter is substantially biased upwards for burst probabilities of 0.25 and higher, when the parameter estimates also become more variable. Recovery of the β parameter is similarly affected by the burst probability. When the burst probability is too high, there are not enough data to obtain reliable parameter estimates. On top of that, for a small burst probability (0.05), there is a tendency to overestimate β . Therefore, we would advise researchers who use our version of the BART (in which p^{burst} is fixed) to only set burst probabilities in the range of 0.1–0.2, as values outside of this range lead to biased parameter estimates. The correlations between γ^+ and β are: −0.55, −0.72, −0.78, −0.83, −0.81, −0.80, −0.83, −0.83, −0.81, −0.77 for burst probabilities of 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45 and “variable”, respectively.

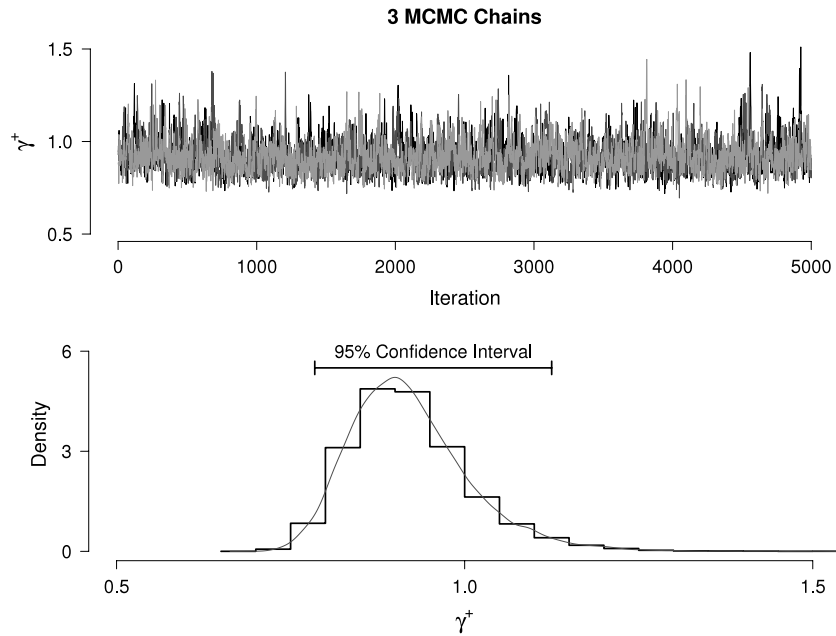


Fig. 2. Top panel: three MCMC chains for parameter γ^+ . Bottom panel: Histogram and non-parametric density estimate for (Silverman, 1986, p. 48) the posterior distribution of parameter γ^+ . The 95% Bayesian confidence interval extends from 0.78 to 1.12.

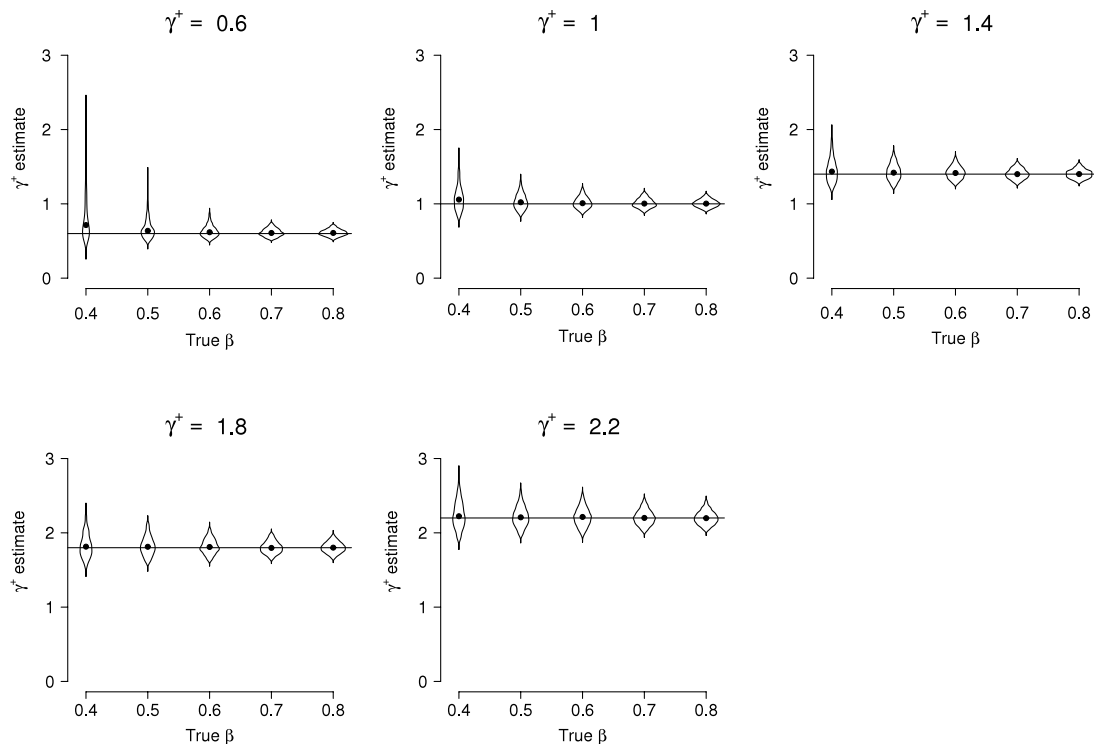


Fig. 3. The 2-parameter BART model recovers parameter γ^+ (results based on a 90-trial BART). Parameter $p^{\text{belief}} = p^{\text{burst}} = 0.15$. The dots represent the median of 1000 posterior means. The violins around the dots are density estimates for the distribution of the 1000 posterior means, with the extreme 5% truncated (see also Hintze & Nelson, 1998). The horizontal lines represent the true parameter values.

In sum, based on the simulation results presented, we conclude that the parameter recovery of the 2-parameter model is very good. In contrast, parameter recovery of the 3- and 4-parameter models is suspect.³ Therefore, we chose to analyze the results from the

empirical study, presented below, with the 2-parameter model only.

4. Experiment

In this section we will present an application of a hierarchical version of the 2-parameter model to empirical BART data. In a within-subjects design, we administered three different doses

³ See the online Appendix at <http://www.donvanravenzwaaij.com>, section "Research & Codes".

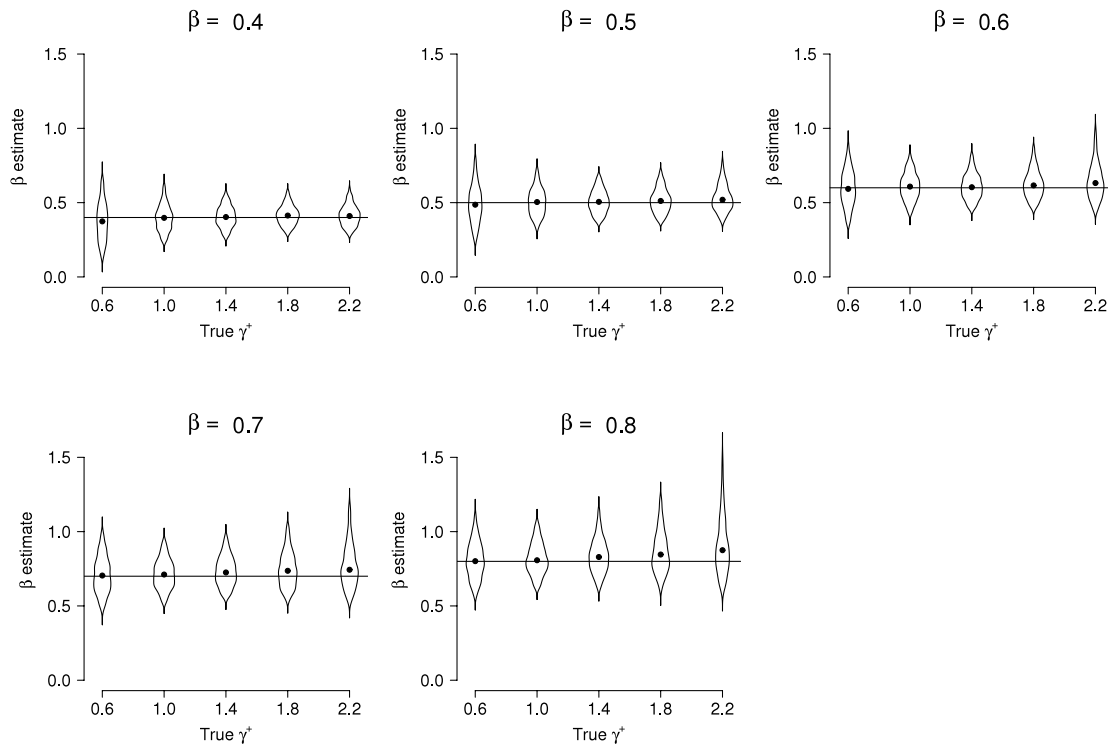


Fig. 4. The 2-parameter BART model recovers parameter β (results based on a 90-trial BART). Parameter $p^{belief} = p^{burst} = 0.15$. The dots represent the median of 1000 posterior means. The violins around the dots are density estimates for the distribution of the 1000 posterior means, with the extreme 5% truncated. The horizontal lines represent the true parameter values.

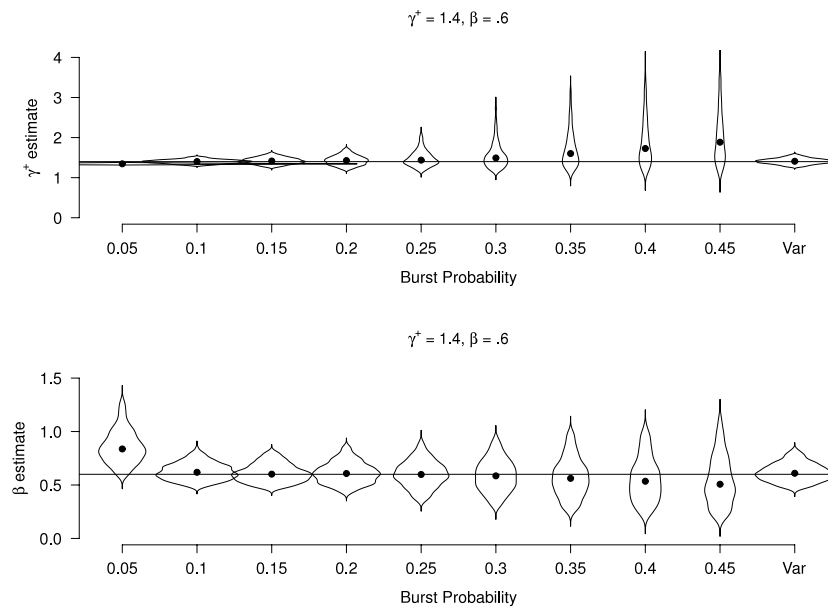


Fig. 5. The 2-parameter BART model recovers parameters γ^+ and β for p^{burst} values from 0.1 to 0.3 (results based on a 90-trial BART). Lower or higher burst probabilities lead to biased estimates. The dots represent the median of 1000 posterior means. The violins around the dots are density estimates for the distribution of the 1000 posterior means, with the extreme 5% truncated. The horizontal lines represent the true parameter values. Var = First 30 trials: burst probability of 0.1, second 30 trials: burst probability of 0.15, third 30 trials: burst probability of 0.2.

of alcohol to every participant, each measured in blood alcohol content, or BAC, in grams per liter: a placebo condition ($BAC = 0$), a tipsy condition ($BAC = 0.5$) and a drunk condition ($BAC = 1$). After consumption, each participant completed a 20 min version of the BART.

We expected that a higher dose of alcohol would lead to more pumps per trial (and therefore a lower percentage of cashing in the experiment). In terms of model parameters, we expected that a higher dose of alcohol would lead to higher risk taking, as evident

in higher values of γ^+ , and to a more diverse, less stable pumping pattern over trials, as evident in lower values of β .

4.1. Method

4.1.1. Participants

Eighteen male students from the University of Amsterdam, aged 18–25, participated in all three conditions in exchange for a monetary reward of 80 euros.

4.1.2. Materials

The amount of alcohol administered to participants was based on Widmark's Formula:

$$BAC = \frac{A}{rW} - \xi t, \quad (6)$$

where BAC is the blood alcohol concentration (in grams per liter), A is the weight of the alcohol consumed since the commencement of drinking (in grams), W is the weight of the person (in kilograms), r is the alcohol distribution ratio (in liters per kilogram), which is on average 0.68 for men, t is the number of hours elapsed since the commencement of drinking, and ξ is the decay factor (Watson, Watson, & Batt, 1981).

Pilot work showed ξ to be approximately 0.15.⁴ Thus, each participant was required to drink an amount of vodka (in milliliters) equal to 1.28 times their body weight (in kilograms) in the tipsy condition and double that amount in the drunk condition. For example, a man weighing 70 kg would be required to drink 90 ml of vodka in the tipsy condition and 180 ml of vodka in the drunk condition.

Participants were required to drink two 0.4 l milkshake cups of fluid. Both cups consisted of half the amount of vodka the participant had to consume, then filled up with multifruit juice. On top of all this, 6 drops of mint oil were added, as earlier tests had shown this to mask both the taste and the scent of the alcohol.

4.1.3. Procedure

Each participant started the experiment at 3 pm. Upon entering, if it was their first session, they received a general instruction about the procedure and signed an informed consent form. Then the participant was asked whether he had drank alcohol the night before, whether he had a light lunch and whether he had consumed any tea, coffee or coke prior that day (the required answers were no, yes, no). Next, the participant got his first breathalyzer measurement. If the BAC read 0 (which it invariably did), the participant was given his first milkshake cup. To finish the first cup the participant was allowed 15 min, after which the second cup was brought in. The participant was allowed 30 min to finish the second cup. After finishing the two cups, the participant received a glass of water and was required to wait for another 20 min for the alcohol to take its full effect. During the complete 65 min, the participant watched a DVD of his choice. After this, a second breathalyzer measurement was obtained. Then the participant completed an unrelated perceptual classification experiment, which took approximately 20 min.⁵ Subsequently, a third breathalyzer measure was obtained. Next, the participant started the BART, which took approximately 20 min to complete. Upon completion of the BART, the participant received a fourth breathalyzer measurement.

4.1.4. Design

In each of the three sessions, a BART was administered with 3 blocks of 30 trials each. In 30 trials the risk associated with the balloon bursting, p^{burst} , was 0.1, in 30 trials the risk was 0.15 and in 30 trials the risk was 0.2 (analogous to p^{burst} = "variable" in the second simulation for the 2-parameter model). The blocks of 30 trials were administered in random order. Parameter p^{burst} did not vary within blocks and was communicated to the participant prior to each block. The amount of money gained with each pump was a percentage of the money accrued so far, chosen such that each pump had an expected gain of exactly zero.

4.2. Results

Fig. 6 shows the within-subject effects for the number of pumps⁶ and, for completeness, the percentage of trials on which the participant decided to cash. Contrary to our expectation, neither alcohol dose nor test session affected these BART performance measures.

4.2.1. Hierarchical Bayesian parameter estimation

We fit the hierarchical 2-parameter model to the data. The model estimates a γ_{ij}^+ and β_{ij} parameter for participants $i = 1, \dots, 18$ for conditions $j = 1, 2, 3$. Our design is a two-way hierarchical Bayesian ANOVA with alcohol dose and session as the independent variables. Because the experimental design did not include all combinations of dose \times session for each participant, dummy variables are used to take only the relevant effects into account (e.g., Ntzoufras, 2009). All 54 parameters are estimated according to

$$\gamma_{ij}^+ = \eta_i^{\gamma^+} + \zeta_{1i}^{\gamma^+} D_{1i} + \zeta_{2i}^{\gamma^+} D_{2i} + \theta_{1i}^{\gamma^+} D_{3i} + \theta_{2i}^{\gamma^+} D_{4i}, \quad (7)$$

$$\beta_{ij} = \eta_i^{\beta} + \zeta_{1i}^{\beta} D_{1i} + \zeta_{2i}^{\beta} D_{2i} + \theta_{1i}^{\beta} D_{3i} + \theta_{2i}^{\beta} D_{4i}, \quad (8)$$

where for both Eqs. (7) and (8) the η_i parameters (the dot indicates the same structure for both equations) is the baseline effect for participant i in the sober condition for the first session. Parameter ζ_{1i} is the alcohol effect from sober to tipsy, parameter ζ_{2i} is the additional alcohol effect from tipsy to drunk. Parameter θ_{1i} is the training effect from session 1 to 2, parameter θ_{2i} is the additional training effect from session 2 to 3. Parameters $D_{\cdot i}$ are dummy variables. For example, to calculate γ^+ for participant 5 in the tipsy condition, which was administered to this participant in session 3, we get

$$\begin{aligned} \gamma_{52}^+ &= \eta_5^{\gamma^+} + \zeta_{15}^{\gamma^+} \times 1 + \zeta_{25}^{\gamma^+} \times 0 + \theta_{15}^{\gamma^+} \times 1 + \theta_{25}^{\gamma^+} \times 1 \\ &= \eta_5^{\gamma^+} + \zeta_{15}^{\gamma^+} + \theta_{15}^{\gamma^+} + \theta_{25}^{\gamma^+}. \end{aligned} \quad (9)$$

For each participant, the effect parameters η_i , $\zeta_{\cdot i}$, and $\theta_{\cdot i}$ are assumed to come from a Gaussian group distribution:

$$\eta_i \sim N(\mu_{\eta}, \sigma_{\eta}^2), \quad (10)$$

$$\zeta_{\cdot i} \sim N(\mu_{\zeta}, \sigma_{\zeta}^2), \quad (11)$$

$$\theta_{\cdot i} \sim N(\mu_{\theta}, \sigma_{\theta}^2). \quad (12)$$

The priors for the mean and the standard deviation of the baseline group Gaussian distributions are given by

$$\mu_{\eta} \sim U(0, 10), \quad (13)$$

$$\sigma_{\eta} \sim U(0, 10). \quad (14)$$

For the alcohol dose and the session effects, we chose to use a prior on effect size instead of on the mean parameter (e.g., Rouder et al., 2009; Wetzels et al., 2009). Thus,

$$\delta_{\cdot \zeta} = \frac{\mu_{\cdot \zeta}}{\sigma_{\cdot \zeta}}, \quad (15)$$

$$\delta_{\cdot \theta} = \frac{\mu_{\cdot \theta}}{\sigma_{\cdot \theta}}. \quad (16)$$

This way one can model the effect size parameters directly, which is convenient if one does not know much about the

⁴ The pilot consisted of administering the beverage to the authors and three other colleagues of the department and measuring their BAC upon consumption.

⁵ The results of this experiment will be published elsewhere.

⁶ Since the mean number of pumps excluding balloons that burst and the mean number of pumps based on all trials do not deviate substantially, we report the mean number of pumps here based on all trials.

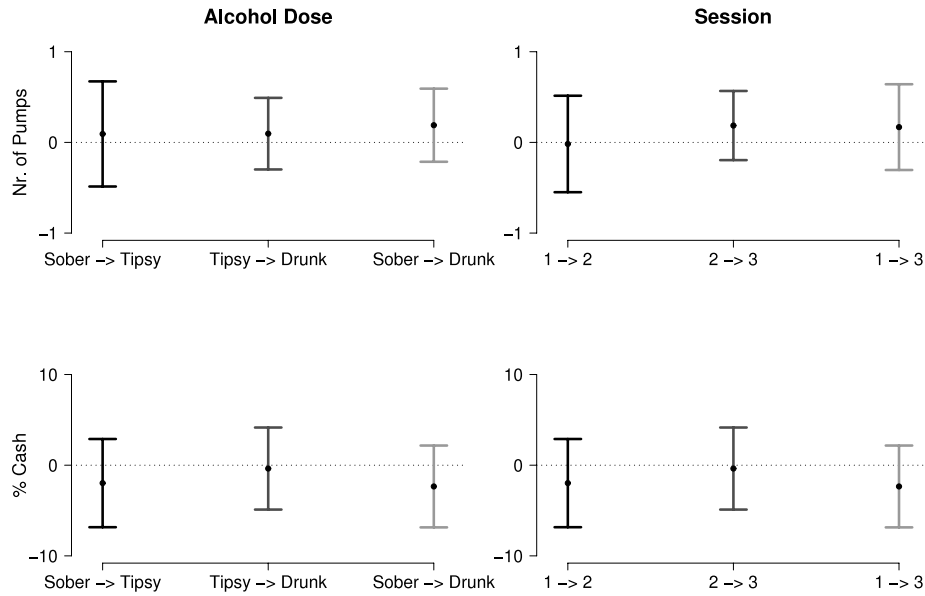


Fig. 6. Alcohol dose (left panels) and test session (right panels) do not affect the mean number of pumps (top panels), nor the percentage of trials on which the participant cashed (bottom panels). Error bars represent 95% frequentist confidence intervals.

underlying scale. The priors for the standard deviations and effect sizes are given by

$$\delta_{\zeta} \sim N(0, 1), \quad (17)$$

$$\sigma_{\zeta} \sim U(0, 10), \quad (18)$$

$$\delta_{\theta} \sim N(0, 1), \quad (19)$$

$$\sigma_{\theta} \sim U(0, 10). \quad (20)$$

For the remaining individual, mean, and standard deviation parameters, we used the following priors: $\gamma_i^+ \sim N(\gamma_{\mu}^+, \gamma_{\sigma}^+)$, $\beta_i \sim N(\beta_{\mu}, \beta_{\sigma})$, $\gamma_{\mu}^+ \sim U(0, 10)$, $\beta_{\mu} \sim U(0, 10)$, $\gamma_{\sigma}^+ \sim U(0, 10)$, $\beta_{\sigma} \sim U(0, 10)$.

For each parameter, we ran three separate Markov chains. For the first chain, we used the following initial values: $\eta_i^{\gamma^+} = \mu_{\eta}^{\gamma^+} = 1.2$, $\eta_i^{\beta} = \mu_{\eta}^{\beta} = 0.5$, $\sigma_{\eta} = \sigma_{\zeta} = \sigma_{\theta} = 1$, and $\zeta_i = \theta_i = \delta_{\zeta} = \delta_{\theta} = 0$. For the second chain, we multiplied all initial values from chain 1 by 0.8, except for $\zeta_i = \theta_i = \delta_{\zeta} = \delta_{\theta}$, which we put on -0.2 . For the third chain, we multiplied all initial values from chain 1 by 1.2, except for $\zeta_i = \theta_i = \delta_{\zeta} = \delta_{\theta}$, which we put on 0.2. We hand-picked initial values based on results from the simulations reported earlier. Each chain consisted of 10,000 iterations, of which the first 5000 were burn-in samples.

4.2.2. Posterior predictives

In Bayesian statistics, model fit can be assessed by means of *posterior predictives*. Posterior predictives are synthetic, model-generated data sets that are produced by parameters drawn from the posterior distribution. If the synthetic data sets closely resemble the empirical data, then the model fit is deemed adequate. We generated posterior predictives for our BART experiment by sampling 1000 values of the parameters γ^+ and β from the joint posterior for each participant and each condition. We then generated 90 trials of BART data with each of the sampled sets of parameters and calculated a mean number of pumps for these 90 trials. Finally, we generated a density across the 1000 sampled mean number of pumps for each participant and each condition. The resulting posterior predictive densities can be seen in Fig. 7. The dots in the figure show the experimental data on the mean number of pumps per participant, with error bars representing 95% frequentist confidence intervals. The vertical

densities next to the data points are the model predictions that follow from the joint posterior. The densities fall smoothly over all confidence intervals, suggesting that the model fits the data well.

4.2.3. Experimental effects

The hierarchical 2-parameter model converged well, with the median of the \hat{R} s over all parameters being 1.01. Fig. 8 displays within-subject effects of the two parameters, γ^+ and β in the left two panels. The γ^+ parameter shows an upward trend with alcohol dosage, and the β parameter shows a downward trend. The confidence intervals, although suggestive, overlap with zero, even for the sober-to-drunk effect. Therefore, we cannot conclude that there is an effect from alcohol consumption on the γ^+ and β parameters. The correlation between the γ^+ and β parameter estimates is $-.68$.

The top right panel displays posterior densities for $\delta_{1\zeta}^{\gamma^+}$ and $\delta_{2\zeta}^{\gamma^+}$. It also displays a posterior density for the effect size from sober to drunk, $\delta_{3\zeta}^{\gamma^+}$, which was estimated in a separate model.⁷ The bottom right panel displays posterior densities for $\delta_{1\zeta}^{\beta}$ and $\delta_{2\zeta}^{\beta}$, as well as a posterior density for the effect size from sober to drunk, $\delta_{3\zeta}^{\beta}$, which was also estimated in a separate model.

To showcase one of the strengths of a hierarchical analysis, we have estimated parameters of the 2-parameter model with maximum likelihood. We have also estimated the parameters Bayesian for each participant and for each condition separately (labeled “individual Bayes”). The left panel of Fig. 9 shows within-subject effects of the maximum likelihood parameter estimates, the middle panel shows within-subject effects of the posterior means of the individual Bayes parameter estimates, and the right panel shows within-subject effects of the posterior means of the hierarchical Bayes parameter estimates (identical to the left panel of Fig. 8). Fig. 9 shows that the 95% confidence intervals are smaller for the hierarchical Bayes model than for both the maximum likelihood and the individual Bayes models. The reason

⁷ Estimating the effect size of the contrasts sober-tipsy and sober-drunk automatically constrains the effect size of the contrast tipsy-drunk. To obtain a posterior for the effect size of the contrast tipsy-drunk, a separate model was necessary to prevent identification problems.

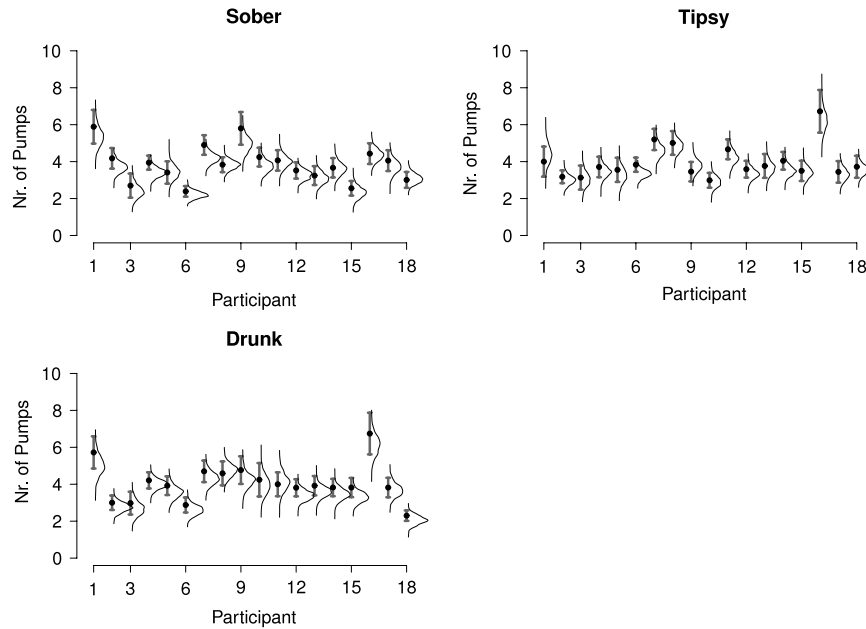


Fig. 7. Posterior predictives indicate that the model fits the data well. Lines: Posterior predictives, based on the model parameter estimates. Dots: experimental effects on the mean number of pumps per participant. Error bars represent 95% confidence intervals of the data.

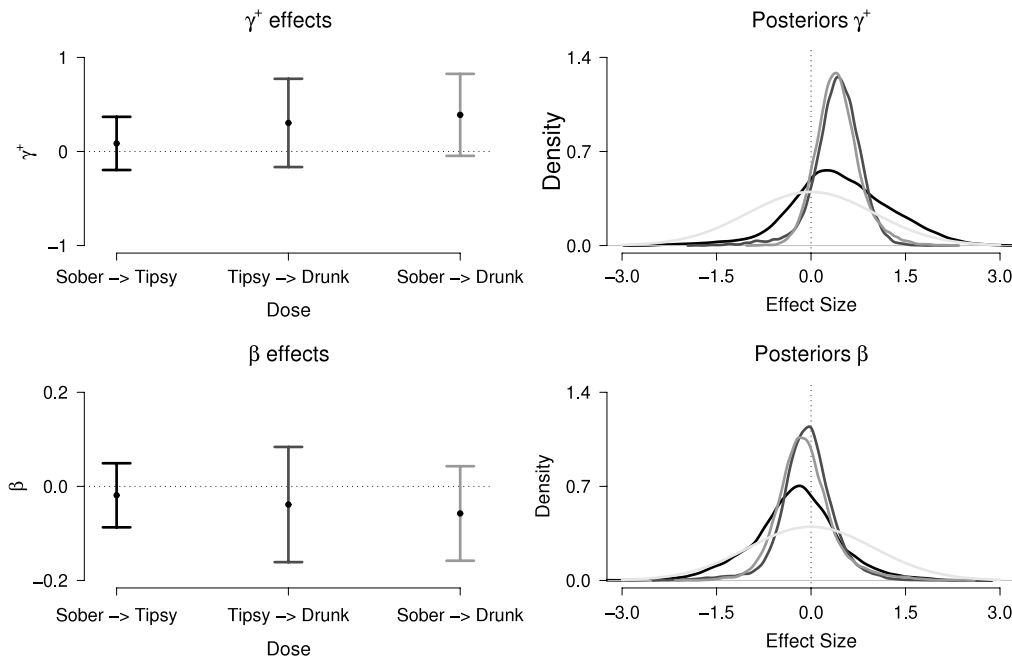


Fig. 8. Parameters γ^+ and β are not clearly affected by alcohol dose. Left Panel: Mean within-subject effects on the γ^+ and β parameters for alcohol dose. Error bars represent 95% confidence intervals. Right Panel: Posterior densities of the effect sizes. Black: $\delta_{1\zeta}$; dark-gray: $\delta_{2\zeta}$; medium-gray: $\delta_{3\zeta}$; light-gray: prior for the effect sizes: $N(0, 1)$.

for this enhanced precision lies in the inclusion of the group-level structure. Since both subject and condition parameters are now drawn from an overarching distribution, parameter estimates will shrink towards the mean. This shrinkage effect is more pronounced for individuals whose parameters are estimated imprecisely, so that the hierarchical Bayesian analysis does not suffer from outliers to the extent that the other two analyses do.

4.2.4. Hierarchical Bayesian hypothesis testing using Bayes factors

So far, all our analyses were concerned with parameter estimation. In this section we will carry out a Bayesian hypothesis test using the *Bayes factor*. Hypothesis testing is important, because

95% confidence intervals cannot quantify evidence in favor of a null-hypothesis that postulates the absence of an effect (Berger & Delampady, 1987; Rouder et al., 2009). In Bayesian testing, every hypothesis that is entertained — here the null-hypothesis H_0 and the alternative hypothesis H_A — is assigned a prior probability. The ratio between two prior model probabilities is known as the *prior odds*, $p(H_0)/p(H_A)$. The prior odds is updated by means of the data and then becomes the *posterior odds*, $p(H_0|D)/p(H_A|D)$. The change from prior odds to posterior odds, $p(D|H_0)/p(D|H_A)$, is known as the *Bayes factor* (Jeffreys, 1961; Kass & Raftery, 1995). Thus,

$$BF_{0A} = \frac{p(H_0|D)}{p(H_A|D)} = \frac{p(D|H_0)}{p(D|H_A)} \times \frac{p(H_0)}{p(H_A)}. \quad (21)$$

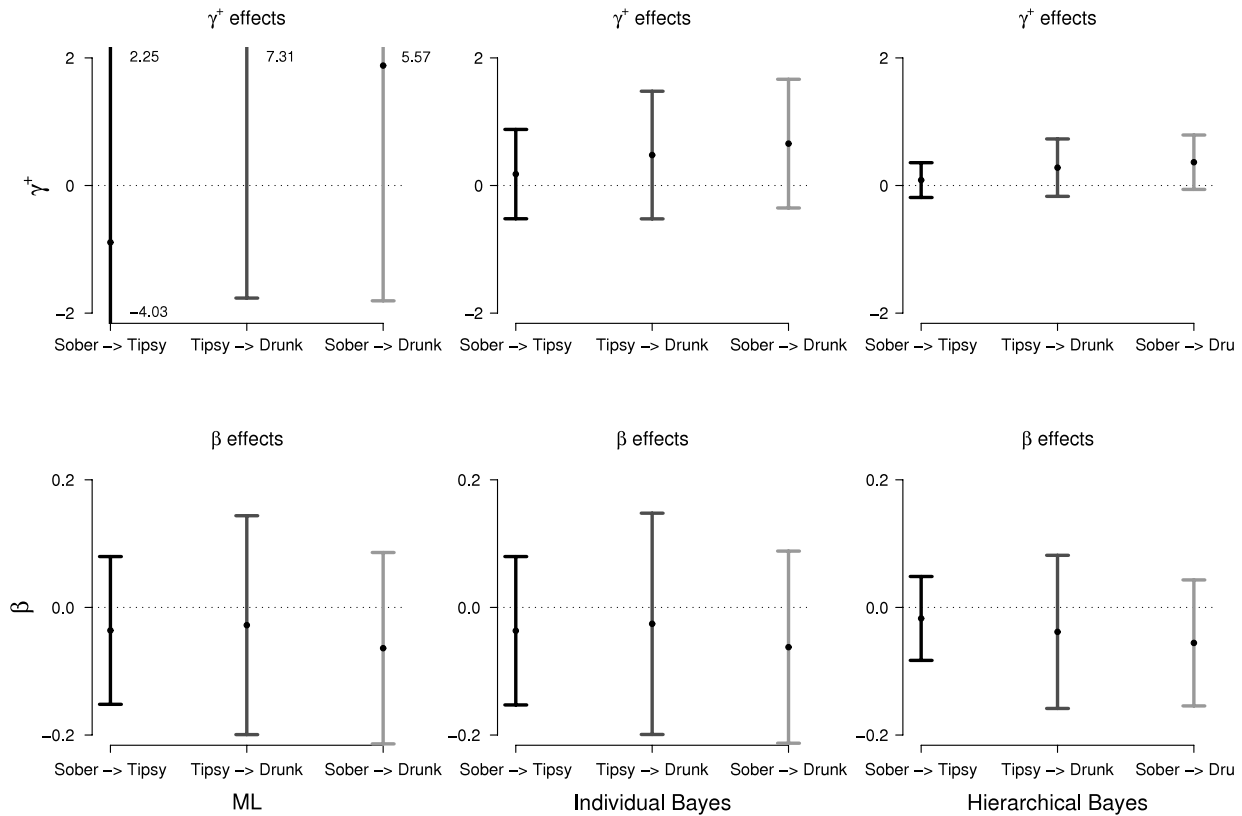


Fig. 9. The 95% frequentist confidence interval of the within-subject effects are smaller for the hierarchical Bayes posterior means (right panel) than for the maximum likelihood parameter estimates (left panel), and the individual Bayes posterior means (middle panel). The small figures in the top left panel indicate the extent of the confidence intervals.

When H_0 and H_A are equally likely a priori, then the Bayes factor is identical to the posterior odds. Bayes factors BF_{0A} higher than 1 indicate support in favor of the null-hypothesis, whereas Bayes factors BF_{0A} lower than 1 indicate support in favor of the alternative hypothesis.

In order to test the null-hypotheses that the effect sizes for each of the BART parameters (i.e., $\delta_{1\zeta}^{\gamma^+}$, $\delta_{2\zeta}^{\gamma^+}$, $\delta_{3\zeta}^{\gamma^+}$, $\delta_{1\zeta}^{\beta}$, $\delta_{2\zeta}^{\beta}$, and $\delta_{3\zeta}^{\beta}$) equals 0, we calculated the Bayes factor for each of the contrasts using the Savage–Dickey method (see e.g. Rouder et al., 2009; Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010; Wetzels et al., 2009). To calculate a Bayes factor using the Savage–Dickey method, one estimates the height of the posterior distribution for the parameter of interest, at the point that is subject to test, and divides this estimate by the height of the prior distribution at that same point.

For the γ^+ parameter, the Bayes factors BF_{0A} for $\delta_{1\zeta}^{\gamma^+}$, $\delta_{2\zeta}^{\gamma^+}$, and $\delta_{3\zeta}^{\gamma^+}$ were estimated to be 1.33, 1.17, and 1.51, respectively. These Bayes factors are inconclusive, supporting the null hypothesis only by the slightest of margins; consequently, we cannot draw any firm conclusions as to whether or not there is an effect from alcohol consumption on the γ^+ parameter. This conclusion echoes the one based on the confidence intervals reported above.

For the β parameter, the Bayes factors BF_{0A} for $\delta_{1\zeta}^{\beta}$, $\delta_{2\zeta}^{\beta}$, and $\delta_{3\zeta}^{\beta}$ were 1.68, 3.04, and 2.61, respectively. These Bayes factors are higher than one, but are still barely worth mentioning, according to the taxonomy by Jeffreys (1961). Therefore, the data are ambiguous with respect to the effect of alcohol on the β parameter.

5. Concluding comments

The first goal of this paper was to increase our knowledge about how the BART models can be applied to empirical data. In order

to do so, we have assessed parameter recovery for a simplified version of the BART model by Wallsten et al. (2005) with 4, 3, and 2 parameters, of which the results for the 4- and 3-parameter versions of the model are reported online. Our second goal was to test the effects of alcohol on performance on the BART task in an experimental setting. We have created a Bayesian hierarchical implementation of the BART model and have applied it to this empirical dataset.

Our simulations indicated that only the 2-parameter model with the risk parameter γ^+ and behavioral consistency parameter β could adequately recover its parameters. The learning parameters present in the full 4-parameter model can not be reliably recovered and even reduce recovery of the other parameters (see also Pleskac, 2008). This suggests that for the specific version of the BART we used with a fixed bursting probability over trials, empirical BART data are not rich enough to warrant the use of more complicated models. Note that even though the learning parameters were statistically unidentified, these parameters are very plausible psychologically. Researchers interested in the learning component of risk taking may better resort to different tasks, such as the Iowa gambling task (Bussemeyer & Stout, 2002; Wetzels et al., 2010).

We also found that the 2-parameter model performs best when the probability of the balloon bursting is in the range of 0.1 to 0.2; researchers interested in applying the BART model to data are advised to use values in this range. A cautionary remark here is in order, as our results were obtained with a BART in which the bursting probability was constant for increasing number of pumps. They may not generalize to the more conventional BART, in which the bursting probability is governed by Eq. (1).

One necessary, but not sufficient, condition that has to be met for every successful measurement model is reliable parameter estimation. We have demonstrated this feature for the 2-parameter model. However, another important condition is

validation (e.g., Vanpaemel, 2009). In order for the BART model to really prove its usefulness, parameter specificity tests would have to be conducted. For instance, the gains for each pump could be increased to see whether this would lead to an increase in the risk taking parameter γ^+ ; or the bursting probability could, unbeknownst to the participant, be modulated between trials to see whether that would lead to more variance in the number of pumps between trials and a decrease in the behavioral consistency parameter β .

In the empirical study, we examined the effects of alcohol consumption on BART performance. In contrast to our expectations, alcohol consumption did not affect the mean number of pumps or the percentage of trials cashed. We then analyzed the behavioral data with a Bayesian hierarchical implementation of the 2-parameter BART model. This analysis showed that alcohol consumption leads to an increase in the risk taking parameter γ^+ , and a decrease in the behavioral consistency parameter β . These effects were, however, relatively modest, and do not allow for clear statistical conclusions. Perhaps our choice for using a BART with a relatively high bursting probability is partially responsible for the somewhat weak effects, as this leads to a relatively limited number of pumps per trial. On the other hand, the hierarchical nature of our design guaranteed that we had plenty of data, so that we may have some confidence in the results of our analyses.

In this paper, we have tried to demonstrate the added value of formal cognitive modeling of tasks used in developmental and clinical psychology. We have also tried to show the benefits of hierarchical Bayesian modeling. Bayesian hierarchical models simultaneously take into account participants' differences and similarities, and they propagate uncertainty and information from different sources in a coherent manner. The growth of cognitive models – Bayesian or otherwise – is increasing (e.g., Johnson, Blaha, Hout, & Townsend, 2010; Maddox, Filoteo, & Zeithamova, 2010; Neufeld, Boksman, Vollick, George, & Carter, 2010; Speekenbrink, Lagnado, Wilkinson, Jahanshahi, & Shanks, 2010; Wenger, Negash, Petersen, & Petersen, 2010), and we hope and expect this trend to continue in the future.

Appendix. WinBUGS code of the BART model

This appendix provides the WinBUGS code for fitting one participant with the 2-parameter BART model. WinBUGS code for the more complicated models is available on the first author's web page, <http://www.donvanravenzwaaij.com/>.

```
model
{
  #Priors for the gamma plus and beta parameters:
  gplus~dunif(0, 10)
  beta~dunif(0, 10)
  for (i in 1:Trials)
  {
    # Amount of pumps DM considers optimal:
    Omega[i] <- -gplus/log(1-pBelief[i])
    for(j in 1:Option[i])
    {
      # Probability that DM will cash for trial i for
      # pump j:
      pCash[i,j] <- 1-(1/(1+exp(beta*(j-Omega[i]))))
      # Choice contains the data as binary variables:
      Choice[i,j]~dbern(pCash[i,j])
    }
  }
}
```

The model requires three variables for input:

- *pbelief* is the participant's belief about the probability that the balloon will burst after any pump.

- *option* is the number of pump opportunities the participant had for each trial.
- *choice* is a binary variable, containing the actual choice for each pump opportunity.

To illustrate, suppose a participant completed a BART with two trials. On the first trial, the participant pumped three times and then cashed. On the second trial, the participant pumped twice and then the balloon burst. The input variables would be represented as follows:

```
option
4 2
choice
0 0 1
0 0
```

References

- Adlaf, E. M., & Smart, R. G. (1983). Risk-taking and drug-use behaviour: an examination. *Drug and Alcohol Dependence*, 11, 287–296.
- Batchelder, W. H. (2007). Cognitive psychometrics: combining two psychological traditions. In *CSCA Lecture*. 2007.
- Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, 275, 1293–1295.
- Berger, J. O., & Delampady, M. (1987). Testing precise hypotheses. *Statistical Science*, 2, 317–352.
- Burian, S. E., Liguori, A., & Robinson, J. H. (2002). Effects of alcohol on risk-taking during simulated driving. *Human Psychopharmacology*, 17, 141–150.
- Bussemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: decomposing performance on the Bechara gambling task. *Psychological Assessment*, 14, 253–262.
- Carlin, B. P., & Louis, T. A. (2000). *Bayes and empirical Bayes methods for data analysis*. New York: Chapman & Hall/CRC.
- Cherpitel, C. J. (1993a). Alcohol and injuries: a review of international emergency room studies. *Addiction*, 88, 651–665.
- Cherpitel, C. J. (1993b). Alcohol, injury, and risk-taking behavior: data from a national sample. *Alcoholism: Clinical and Experimental Research*, 17, 762–766.
- Cohen, J., Dearnaley, E. J., & Hansel, C. E. M. (1958). The risk undertaken in driving under the influence of alcohol. *British Medical Journal*, 1, 1438–1442.
- Congdon, P. D. (2010). *Applied Bayesian hierarchical methods*. Boca Raton: CRC Press.
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. Ft Worth: Dryden Press.
- Edwards, A. W. F. (1992). *Likelihood*. Baltimore, MD: The Johns Hopkins University Press.
- Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, 53, 134–140.
- Estes, W. K. (2002). Traps in the route to models of memory and decision. *Psychonomic Bulletin & Review*, 9, 3–25.
- Eysenck, S. B., & Eysenck, H. J. (1977). The place of impulsiveness in a dimensional system of personality description. *British Journal of Social Clinical Psychology*, 16, 57–68.
- Fromme, K., Katz, E., & D'Amico, E. (1997). Effects of alcohol intoxication on the perceived consequences of risk taking. *Experimental Clinical Psychopharmacology*, 5, 14–23.
- Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review*, 116, 439–453.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457–472.
- Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law revealed: the case for an exponential law of practice. *Psychonomic Bulletin & Review*, 7, 185–207.
- Hintze, J., & Nelson, R. (1998). Violin plots: a box plot-density trace synergism. *The American Statistician*, 52(2).
- Hopko, D., Lejuez, C., Daughters, S., Aklon, W., Osborne, A., & Simmons, B. (2005). Construct validity of the Balloon Analogue Risk Task (BART): relationship with MDMA use by inner-city drug users in residential treatment. *Journal of Psychopathology and Behavioral Assessment*, 28, 95–101.
- Jeffreys, H. (1961). *Theory of probability*. Oxford, UK: Oxford University Press.
- Johnson, S. A., Blaha, L. M., Hout, J. W., & Townsend, J. T. (2010). Systems factorial technology provides new insights on global-local information processing in autism spectrum disorders. *Journal of Mathematical Psychology*, 54, 53–72.
- Kalichman, S. C., Heckman, T., & Kelly, J. A. (1996). Sensation seeking as an explanation for the association between substance use and HIV-related risky sexual behavior. *Archives of Sexual Behavior*, 25, 141–154.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773–795.

- Lee, M. D., & Wagenmakers, E.-J. (2009). A course in Bayesian graphical modeling for cognitive science. Course notes available at <http://users.fmg.uva.nl/ewagenmakers/BayesCourse/BayesBook.pdf>.
- Lee, M. D., & Webb, M. R. (2005). Modeling individual differences in cognition. *Psychonomic Bulletin & Review*, 12, 605–621.
- Leigh, B. C., & Stall, R. (1993). Substance use and risky sexual behavior for exposure to HIV: issues in methodology, interpretation, and prevention. *American Psychologist*, 48, 1035–1045.
- Lejuez, C. W., Aklin, W. M., Zvolensky, M. J., & Pedulla, C. M. (2003). Evaluation of the balloon analogue risk task (BART) as a predictor of adolescent real-world risk-taking behaviors. *Journal of Adolescence*, 26, 475–479.
- Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., & Stuart, G. L. (2002). Evaluation of a behavioral measure of risk taking: the balloon analogue risk task (BART). *Journal of Experimental Psychology: Applied*, 8, 75–84.
- 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.
- Maddox, W. T., Filoteo, J. V., & Zeithamova, D. (2010). Computational models inform clinical science and assessment: an application to category learning in striatal-damaged patients. *Journal of Mathematical Psychology*, 54, 109–122.
- McEwan, R. T., McCallum, A., Bhopal, R. S., & Madhok, R. (1992). Sex and the risk of HIV infection: the role of alcohol. *British Journal of Addiction*, 87, 577–584.
- Merkle, E., Smithson, M., & Verkuilen, J. (2011). Hierarchical models of simple mechanisms underlying confidence in decision making. *Journal of Mathematical Psychology*, 55(1), 57–67.
- Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47, 90–100.
- Neufeld, R. W. J., Boksman, K., Vollick, D., George, L., & Carter, J. R. (2010). Stochastic dynamics of stimulus encoding in schizophrenia: theory, testing, and application. *Journal of Mathematical Psychology*, 54, 90–108.
- Nilsson, H., Rieskamp, J., & Wagenmakers, E.-J. (2011). Hierarchical Bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology*, 55(1), 84–93.
- Ntzoufras, I. (2009). *Bayesian modeling using WinBUGS*. Hoboken: Wiley.
- Pleskac, T. J. (2008). Decision making and learning while taking sequential risks. *Journal of Experimental Psychology – Learning Memory and Cognition*, 34, 167–185.
- Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, 12, 573–604.
- Rouder, J. N., Lu, J., Morey, R. D., Sun, D., & Speckman, P. L. (2008). A hierarchical process dissociation model. *Journal of Experimental Psychology: General*, 137, 370–389.
- Rouder, J. N., Lu, J., Speckman, P. L., Sun, D., & Jiang, Y. (2005). A hierarchical model for estimating response time distributions. *Psychonomic Bulletin & Review*, 12, 195–223.
- Rouder, J. N., Lu, J., Sun, D., Speckman, P., Morey, R., & Naveh-Benjamin, M. (2007). Signal detection models with random participant and item effects. *Psychometrika*, 72, 621–642.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t*-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16, 225–237.
- Sheu, C.-F., & O'Curry, S. L. (1998). Simulation-based Bayesian inference using BUGS. *Behavioral Research Methods, Instruments, & Computers*, 30, 232–237.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman & Hall.
- Speekenbrink, M., Lagnado, D. A., Wilkinson, L., Jahanshahi, M., & Shanks, D. R. (2010). Models of probabilistic category learning in Parkinson's disease: strategy use and the effects of *L*-dopa. *Journal of Mathematical Psychology*, 54, 123–136.
- Vanpaemel, W. (2009). BayesGCM: software for Bayesian inference with the generalized context model. *Behavior Research Methods*, 41, 1111–1120.
- Wagenmakers, E.-J., Lee, M. D., Lodewyckx, T., & Iverson, G. (2008). Bayesian versus frequentist inference. In H. Hoijtink, I. Klugkistand, & P. A. Boelen (Eds.), *Bayesian evaluation of informative hypotheses* (pp. 181–207). New York: Springer.
- Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. P. P. P. (2010). Bayesian hypothesis testing for psychologists: a tutorial on the Savage–Dickey method. *Cognitive Psychology*, 60, 158–159.
- Wallach, M. A., Kogan, N., & Bem, D. J. (1962). Group influence on individual risk taking. *Journal of Abnormal and Social Psychology*, 65, 75–86.
- Wallsten, T. S., Pleskac, T. J., & Lejuez, C. W. (2005). Modeling behavior in a clinically diagnostic sequential risk-taking task. *Psychological Review*, 112, 862–880.
- Watson, P. E., Watson, I. D., & Batt, R. D. (1981). Prediction of blood alcohol concentrations in human subjects. *Journal of Studies on Alcohol*, 42, 547–556.
- Wenger, M. J., Negash, S., Petersen, R. C., & Petersen, L. (2010). Modeling and estimating recall processing capacity: sensitivity and diagnostic utility in application to mild cognitive impairment. *Journal of Mathematical Psychology*, 54, 73–89.
- Wetzels, R., Raaijmakers, J. G. W., Jakab, E., & Wagenmakers, E.-J. (2009). How to quantify support for and against the null hypothesis: a flexible WinBUGS implementation of a default Bayesian *t*-test. *Psychonomic Bulletin & Review*, 16, 752–760.
- Wetzels, R., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E.-J. (2010). Bayesian parameter estimation in the expectancy valence model of the Iowa gambling task. *Journal of Mathematical Psychology*, 54, 14–27.