



# Probabilistic Inference: Task Dependency and Individual Differences of Probability Weighting Revealed by Hierarchical Bayesian Modeling

Moritz Boos<sup>†</sup>, Caroline Seer, Florian Lange and Bruno Kopp<sup>\*</sup>

Department of Neurology, Hannover Medical School, Hannover, Germany

## OPEN ACCESS

### Edited by:

Pietro Cipresso,  
IRCCS Istituto Auxologico Italiano, Italy

### Reviewed by:

Richard S. John,  
University of Southern California, USA  
Martin Lages,  
University of Glasgow, UK

### \*Correspondence:

Bruno Kopp  
kopp.bruno@mh-hannover.de

### † Present Address:

Moritz Boos,  
Department of Psychology, Carl von  
Ossietzky University Oldenburg,  
Oldenburg, Germany

### Specialty section:

This article was submitted to  
Quantitative Psychology and  
Measurement,  
a section of the journal  
Frontiers in Psychology

Received: 06 January 2016

Accepted: 06 May 2016

Published: 27 May 2016

### Citation:

Boos M, Seer C, Lange F and Kopp B  
(2016) Probabilistic Inference: Task  
Dependency and Individual  
Differences of Probability Weighting  
Revealed by Hierarchical Bayesian  
Modeling. *Front. Psychol.* 7:755.  
doi: 10.3389/fpsyg.2016.00755

Cognitive determinants of probabilistic inference were examined using hierarchical Bayesian modeling techniques. A classic urn-ball paradigm served as experimental strategy, involving a factorial two (prior probabilities) by two (likelihoods) design. Five computational models of cognitive processes were compared with the observed behavior. Parameter-free Bayesian posterior probabilities and parameter-free base rate neglect provided inadequate models of probabilistic inference. The introduction of distorted subjective probabilities yielded more robust and generalizable results. A general class of (inverted) S-shaped probability weighting functions had been proposed; however, the possibility of large differences in probability distortions not only across experimental conditions, but also across individuals, seems critical for the model's success. It also seems advantageous to consider individual differences in parameters of probability weighting as being sampled from weakly informative prior distributions of individual parameter values. Thus, the results from hierarchical Bayesian modeling converge with previous results in revealing that probability weighting parameters show considerable task dependency and individual differences. Methodologically, this work exemplifies the usefulness of hierarchical Bayesian modeling techniques for cognitive psychology. Theoretically, human probabilistic inference might be best described as the application of individualized strategic policies for Bayesian belief revision.

**Keywords:** hierarchical Bayesian modeling, probabilistic inference, Bayesian inference, probability weighting, prospect theory

## INTRODUCTION

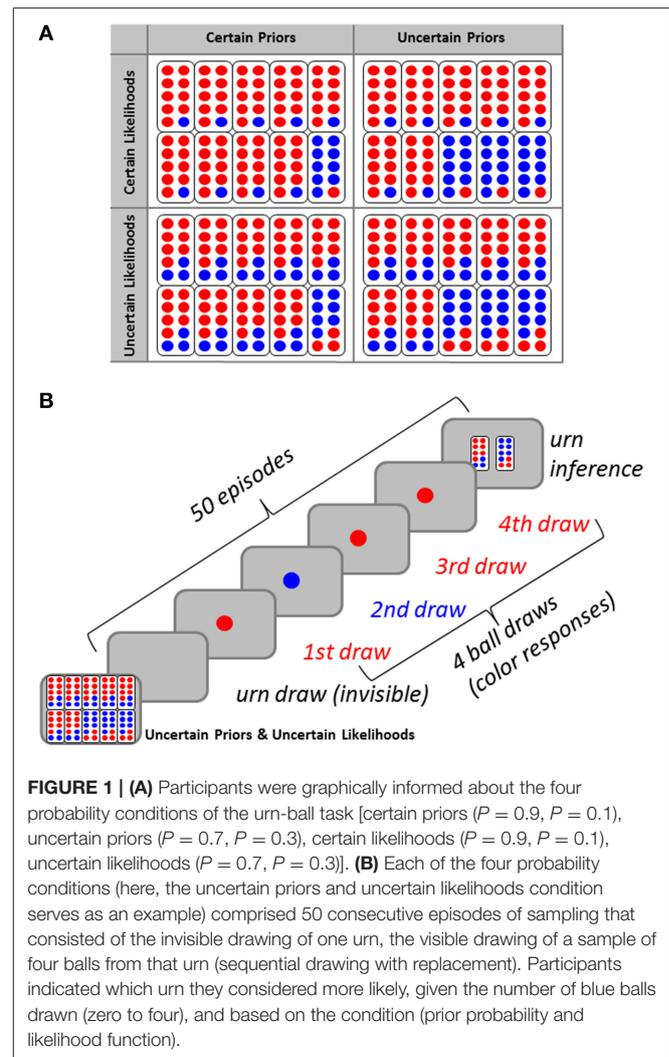
Knight (1921) distinguished between risky worlds, referring to situations where perfect knowledge about probabilities is present and uncertain worlds, referring to situations where probabilities remain unknown. Savage (1954) made a similar distinction when he introduced the term *small worlds* for situations where all alternatives and their probabilities are known. In contrast, relevant information is unknown and/or must be estimated in *large worlds* (see also Johnson and Busemeyer, 2010; Volz and Gigerenzer, 2012).

Small worlds provide opportunities for analysing Bayesian inference. In Bayesian decision theory (Jaynes, 2003; MacKay, 2003; Robert, 2007), degrees of belief in states of the world are specified. Bayesian inference updates prior beliefs using new evidence to derive posterior beliefs. **Figure 1A** shows small worlds, vested as an urn-ball task (Phillips and Edwards, 1966), consisting of two binary random variables, one representing unobservable states of the world (i.e., urns,  $H \in \{H_1, H_2\}$ ), the other representing observable events (i.e., balls,  $e \in \{0, 1\}$ ). Participants were asked for inferences about the current hidden state of that world, given small samples of events which could have been generated from either state (**Figure 1B**). To introduce experimental variance, we manipulated the task's probabilistic contingencies at two levels. First, we introduced uncertainty about the sort of urn containing balls (by sampling urns from two probability distributions). Second, we manipulated the proportion of ball colors within each urn. We will refer to these probabilistic contingencies as prior probabilities and likelihoods, respectively.

Thus, participants are informed about the world's stochastic structure, and they have access to evidence generated from one of these states. Participants represent knowledge-based and evidence-based degrees of belief in world states as probabilities. Specifically, participants hold informed prior beliefs [i.e.,  $P(H_1)$ ,  $P(H_2) = 1 - P(H_1)$ ] and likelihoods [i.e.,  $P(E|H_1)$ ,  $P(E|H_2)$ ], i.e., conditional probabilities for some evidence, given each of the states. According to Bayes' theorem, prior probabilities are combined with likelihoods to provide posterior probabilities (Gold and Shadlen, 2007).

A long history of studies demonstrates that human judgment deviates from Bayesian decision theory (Kahneman et al., 1982). Initially, Edwards coined the term "conservatism" to describe probabilistic inference in which persons overweigh prior beliefs (base rates) and underweigh new sample evidence when compared to Bayesian decision theory (Edwards, 1982). Shanteau's work (1972, 1975) was centered around delineating conditions for the cognitive (sub-)additivity of new sample evidence. Later research identified the *base rate neglect* (Kahneman and Tversky, 1973; Bar-Hillel, 1980), i.e., a cognitive bias which indicates that the posterior probability of hypothesis  $H$ , given evidence  $e$ , is assessed without taking into account the prior probability (*base rate*) of  $H$ . Base rate neglect represents a particularly pertinent class of deviations in probabilistic judgment from Bayesian decision theory (Koehler, 1996).

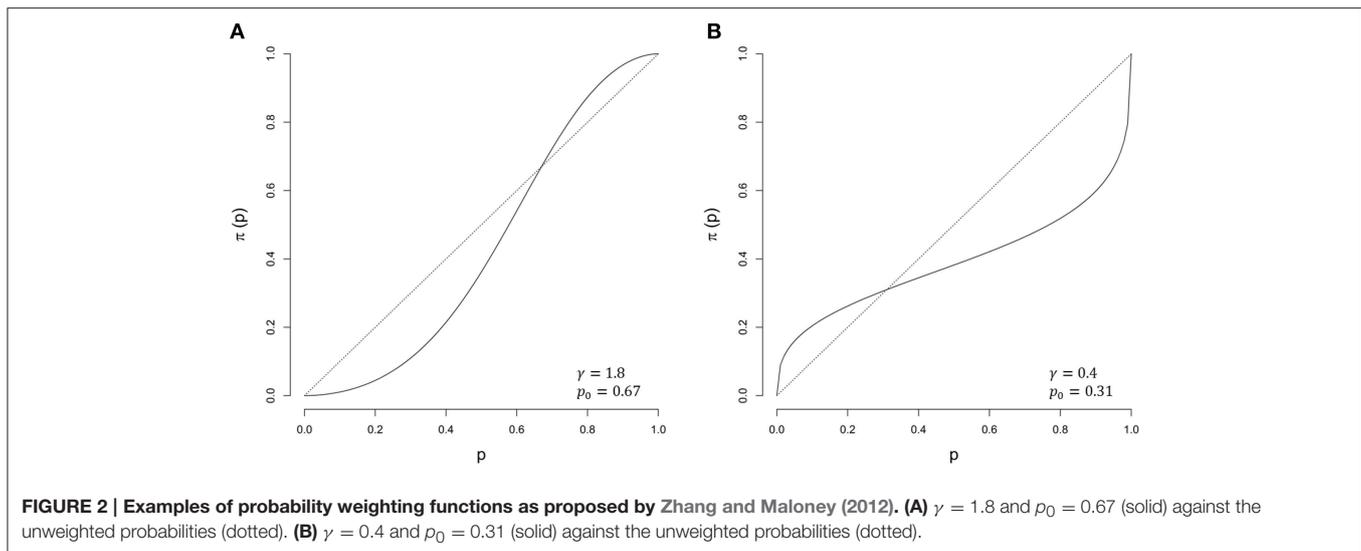
Prospect Theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992) successfully describes economic decision behavior. Its probabilistic part proposed subjective probabilities used in decision-making to be non-linear functions of objective probabilities, with their relationships being best described by S-shaped probability weighting functions, leading to a tendency to overestimate low probabilities and to underestimate high probabilities (see **Figure 2**). One of our reviewers brought our attention to the fact that the notion of weighting probabilities was around 15 years prior to the introduction of Prospect Theory (Edwards, 1954), and that Edwards' early discussion of probability weighting is in the context of probabilistic inference rather than risky choice, which is the focus of Prospect Theory.



**FIGURE 1 | (A)** Participants were graphically informed about the four probability conditions of the urn-ball task [certain priors ( $P = 0.9$ ,  $P = 0.1$ ), uncertain priors ( $P = 0.7$ ,  $P = 0.3$ ), certain likelihoods ( $P = 0.9$ ,  $P = 0.1$ ), uncertain likelihoods ( $P = 0.7$ ,  $P = 0.3$ )]. **(B)** Each of the four probability conditions (here, the uncertain priors and uncertain likelihoods condition serves as an example) comprised 50 consecutive episodes of sampling that consisted of the invisible drawing of one urn, the visible drawing of a sample of four balls from that urn (sequential drawing with replacement). Participants indicated which urn they considered more likely, given the number of blue balls drawn (zero to four), and based on the condition (prior probability and likelihood function).

Such S-shaped probability distortions are ubiquitous in research on probabilistic inference (Gonzalez and Wu, 1999; Luce, 2000; Zhang and Maloney, 2012; Cavagnaro et al., 2013). Stott (2006) provides a review of various probability weighting functions and a summary of parameter values reported in the literature for each of the weighting functions. We refer the interested reader to this publication for descriptions of multiple weighting functions. Earlier research on probability weighting functions revealed considerable inter- and intra-individual variance of probability weighting parameters (Tversky and Kahneman, 1992; Gonzalez and Wu, 1999; Stott, 2006; Wu et al., 2009), but no former study addressed the question whether Bayesian inference in a risky/small world involves probability weighting through Bayesian modeling approaches.

The present study aimed at contributing to the literature by applying Bayesian statistics (Jaynes, 2003; MacKay, 2003; Robert, 2007) to model Bayesian inference. Bayesian methods have become increasingly accepted for data analysis in cognitive science (Edwards et al., 1963; Wagenmakers, 2007; Gallistel, 2009; Kruschke, 2010; Lee, 2011; Hoijtink, 2012).



Specifically, hierarchical Bayesian modeling provides flexible and interpretable ways of analysing formal models of cognitive processes (Lee, 2011). Hierarchical Bayesian models are models in which parameters are sampled from distributions determined by other parameters (so-called hyper-parameters; Shiffrin et al., 2008). By making the individual parameters dependent on their group mean, a trade-off between fitting the group as a whole and fitting each individual separately is introduced (Shiffrin et al., 2008), allowing for improved predictive robustness of the model (Gelman et al., 2014a). Here, we applied hierarchical Bayesian modeling to Bayesian inference in order to examine whether human subjects (a) follow normative Bayesian specifications, and (b) are influenced by non-normative tendencies, such as base rate neglect or (inverse) S-shaped probability weighting. Notice that this is the first study that applied hierarchical Bayesian modeling to Bayesian inference in a risky/small world.

## MATERIALS AND METHODS

### Participants

Sixteen psychology students ( $I = 16$ ) participated for course credit (15 female, 1 male). Age ranged from 19 to 50 years ( $M = 24.7$ ;  $SD = 9.3$ ). All participants indicated having normal or corrected-to-normal sight. The study was reviewed and approved by the Ethics Committee of TU Braunschweig (Department of Life Sciences). Informed consent was obtained from all subjects.

### Inference Task

The inference task is a modification of tasks used in Phillips and Edwards (1966) and Grether (1980, 1992; see also Stern et al., 2010; Achtziger et al., 2014). Factorial combination of two levels of prior probabilities and two levels of the likelihoods yielded four experimental conditions which were administered to each participant. Their order was counterbalanced across participants under the following restriction: Prior probability was the slowly varying factor (one level of prior probabilities was repeated across

two successive blocks of trials), whereas the likelihoods changed from block to block, yielding four different orders.

At the beginning of each condition, two types of urns (labeled  $H_1$  and  $H_2$  here, respectively) were presented on a computer screen. In uncertain likelihood conditions ( $L_u$ ), urn type  $H_1$  contained seven blue ( $e = 1$ ) and three red ( $e = 0$ ) balls [i.e.,  $P(e = 1|H_1) = 0.7$ ,  $P(e = 0|H_1) = 0.3$ ], while urn type  $H_2$  contained three blue and seven red balls [i.e.,  $P(e = 1|H_2) = 0.3$ ,  $P(e = 0|H_2) = 0.7$ ]. In certain likelihood conditions ( $L_c$ ), urn type  $H_1$  contained nine blue balls and one red ball [i.e.,  $P(e = 1|H_1) = 0.9$ ,  $P(e = 0|H_1) = 0.1$ ], while urn type  $H_2$  contained one blue ball and nine red balls [i.e.,  $P(e = 1|H_2) = 0.1$ ,  $P(e = 0|H_2) = 0.9$ ]. Prior probabilities were manipulated by presenting 10 urns, composed of different numbers of type  $H_1$  and type  $H_2$  urns. In uncertain prior probability conditions ( $P_u$ ), three type  $H_1$  and seven type  $H_2$  urns [i.e.,  $P(H_1) = 0.3$ ,  $P(H_2) = 0.7$ ] were present. In the certain prior probability condition ( $P_c$ ), one type  $H_1$  urn and nine type  $H_2$  urns [i.e.,  $P(H_1) = 0.1$ ,  $P(H_2) = 0.9$ ] were present. In the absence of a previous study that examined human probabilistic inference via hierarchical Bayesian modeling (see below), these quite unbalanced prior probabilities and likelihoods were chosen to minimize a potential role of pure “guessing.” It should be noted that the results may have been biased by the selection of these parameters. Colors were counterbalanced across participants, but we will ignore this in our description to avoid confusion.

Each experimental trial (Figure 1B) consisted of the following sequence of events: First, one urn was selected randomly, with the outcome of this selection remaining unknown to the participant. Subsequently, a random sample of four balls ( $K = 4$ ) was drawn sequentially with replacement from that urn, shown one by one. Ignoring the order of drawn balls, this procedure generated five distinct possible situations (i.e., zero to four blue balls) in each condition. Ball stimuli were presented in the center of a computer screen (Eizo FlexScan T766 19”; Hakusan, Ishikawa, Japan) against gray background (size = one degree, duration = 100 ms, stimulus onset asynchrony = 2500 ms). Trial duration

amounted to around 10 s. Participants were asked to indicate the color of each stimulus by pressing the left or right Ctrl key on a standard computer keyboard. Participants indicated which urn they considered more likely, given the number of blue balls drawn (zero to four), and based on the condition (prior probability and likelihood function; Data Sheet 1 in Supplementary Material). They indicated their inference by pressing the left or right Ctrl key for urn type  $H_1$  and urn type  $H_2$ , respectively. Feedback was not provided in order to avoid confounding of probabilistic inference proper and evaluative processes, pending two-factor models of decision-making such as, for example, Prospect Theory (Kahneman and Tversky, 1979), although Grether and Plott (1979) have demonstrated that incentives may have little influence over performance in probabilistic inference tasks. Stimulus-response mapping was counterbalanced across participants. Neither feedback nor reward was provided during the experiment.

Each participant completed four practice trials under supervision of the experimenter, each with one urn type  $H_1$  and one urn type  $H_2$  exemplar, to become accustomed to the task. Successful completion of the practice trials demonstrated that participants understood the procedure and their task. There were  $N = 50$  trials per condition, with short breaks between the conditions. We chose 50 trials per condition because the study was also designed to measure event-related potentials (Seer et al., 2016; Kopp et al., under review), and event-related potentials require large numbers of trials for averaging. Given five possible outcomes per condition, participants responded to each possible outcome an average of 10 times per condition. Thus, it should be mentioned that the responding had a quite repetitive character, and that participants may have at times simply recalled their responses from earlier trials in later trials. The experiment was run using Presentation<sup>®</sup> (Neurobehavioral Systems, Albany, CA).

## Bayesian Inference

Binary inferences were requested (Figure 1). Given prior probabilities  $P(H_1) = 1 - P(H_2)$  for two hypothetical states  $H_1$  and  $H_2$ , and a set of binary data  $E = \{e_1, \dots, e_K\}$  with  $K \in \mathbb{N}$ , we can compute the posterior probabilities as

$$P(H_1 | E) = \frac{P(H_1) \cdot P(E | H_1)}{P(E)}, \quad (1)$$

and

$$P(H_2 | E) = \frac{P(H_2) \cdot P(E | H_2)}{P(E)}, \quad (2)$$

where  $P(H_1|E) = 1 - P(H_2|E)$ . Formulating these posterior probabilities in log-odds form (Jaynes, 2003), using the notation  $Lo(P(H_1|E))$  for the posterior log-odds favoring  $H_1$  we obtain,

$$Lo(P(H_1|E)) = \ln \frac{P(H_1) \cdot P(E | H_1)}{P(H_2) \cdot P(E | H_2)}. \quad (3)$$

For each single binary datum  $e_k$ , with  $k \in [1, \dots, K]$ , we obtain,

$$Lo(P(H_1|E)) = \ln \frac{P(H_1)}{P(H_2)} + \sum_{k=1}^K \ln \frac{P(e_k | H_1)}{P(e_k | H_2)}. \quad (4)$$

Hence, the Bayesian updating in log-odds form for a binary hypothesis equals adding the logarithm of the likelihood ratio  $\ln P(e_k|H_1)/P(e_k|H_2)$  to the logarithm of the prior odds  $\ln P(H_1)/P(H_2)$ . We can therefore represent the accumulation of evidence by adding a new likelihood ratio for each new datum  $e_k$ .

## Probability Weighting

Zhang and Maloney (2012) proposed a linear probability weighting function in the log-odds space, capable of modeling probability weighting as in Prospect Theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). We chose this probability weighting function for its compatibility with the idea of Bayesian updating (see Section Bayesian Inference with Weighted Probabilities). Zhang and Maloney (2012) used the formula

$$\pi = \text{logistic}(\gamma \cdot Lo(p) + (1 - \gamma) \cdot Lo(p_0)), \quad (5)$$

with  $p$  being a probability and the two weighting-parameters  $\gamma \in [0, \infty)$  and  $p_0 \in [0, 1]$ ,  $\text{logistic}(x) = \frac{1}{1 + e^{-x}}$  being the logistic function,  $\pi$  being the corresponding weighted probability, and  $Lo(p)$  given by

$$Lo(p) = \ln \frac{p}{1-p}. \quad (6)$$

Note that,  $\text{logistic}(Lo(p)) = p$ .

The unknown parameters are the slope of the weighting function  $\gamma$  and the point  $p_0$ , where the weighted probability equals the unweighted probability,

$$Lo(\pi(p_0)) = Lo(p_0). \quad (7)$$

The slope parameter  $\gamma$  determines the shape of the weighting function. If  $\gamma = 1$ , the probabilities remain untransformed. If  $\gamma < 1$ , the weighting function has an inverse-S shape or a concave shape if  $p_0$  approaches one. If  $\gamma > 1$ , the weighting function shows an S-shape or a convex shape if  $p_0$  approaches one (Figure 2). Using a comparable log-odds approach, Zhang and Maloney (2012) successfully modeled choice and confidence data from a range of studies.

## Bayesian Inference with Weighted Probabilities

Since Zhang and Maloney (2012) provide a linear weighting function in log-odds space, we can easily represent Bayesian inference with transformed log-odds. Applying Equation (5) to the log-odds evidence  $Lo(P(H_1|E))$  as defined in Equation (3) we obtain

$$Lo(\pi(H_1|E)) = \gamma \cdot \ln \frac{P(H_1)}{P(H_2)} + \gamma \cdot \ln \frac{P(E | H_1)}{P(E | H_2)} + (1 - \gamma) \cdot Lo(p_0). \quad (8)$$

Therefore, we can also represent the weighted accumulation of evidence as the addition of

$$\gamma \cdot \ln \frac{P(E | H_1)}{P(E | H_2)} = \sum_{k=1}^K \gamma \cdot \ln \frac{P(e_k | H_1)}{P(e_k | H_2)}, \quad (9)$$

for each single datum  $e_k$ .

## Models

### Unweighted Bayesian Posterior Probabilities

The simplest model incorporating Bayesian inference relies on unweighted posterior probabilities. This model is Bayes optimal, no other classifier of the urn state can do better on average. We use the posterior to predict the inference of  $H_1$ , on any trial, using the sampling statement

$$y_n \sim \text{bern}(P(H_1|E)). \tag{10}$$

Here,  $y_n$  represents the  $n$ -th inference ( $y_n \in \{0, 1\}$  with  $n \in \{1, \dots, N\}$ ). A value of zero denotes the inference of the urn with higher prior probability and a value of one the inference of the urn with lower prior probability.  $P(H_1|E)$  is the  $n$ -th posterior probability for  $H_1$ , given a sample of evidence  $E$ , as defined in Equation (1).  $\text{bern}(P(H_1|E))$  denotes the Bernoulli distribution:  $y_n$  takes on value one with probability  $P(H_1|E)$ , and  $y_n$  equals zero with probability  $1 - P(H_1|E)$ . There are five different possibilities to draw combinations of blue and red balls (disregarding sequential order); therefore there are five different values of  $P(H_1|E)$  and  $Lo(P(H_1|E))$  (Equation 3) per experimental condition which are shown in **Table 1**.

### Base Rate Neglect

Base rate neglect leads to the assumption that  $\pi(P(H_1)) = \pi(P(H_2)) = 0.5$ , irrespective of the veridical prior probabilities  $P(H_1)$  and  $P(H_2)$ . The model of Bayesian inference thus reduces to the likelihood of the data, i.e., to

$$y_n \sim \text{bern}(P(E|H_1)). \tag{11}$$

### Weighted Bayesian Posterior Probabilities

In the following, we describe three models of weighted Bayesian posterior probabilities (Equation 8).

**TABLE 1 | Posterior probability for the rare urn  $P(H_1|E)$  and  $Lo(P(H_1|E))$  for different conditions and different ball samples drawn.**

Number of blue balls drawn	Condition			
	$P_{cLc}$	$P_{uLc}$	$P_{cLu}$	$P_{uLu}$
<b><math>P(H_1 E)</math></b>				
Zero out of four	0.00002	0.00007	0.00373	0.01425
One out of four	0.00137	0.01526	0.02	0.07297
Two out of four	0.1	0.3	0.1	0.3
Three out of four	0.9	0.972	0.37692	0.7
Four out of four	0.99863	0.99964	0.76709	0.92703
<b><math>Lo(P(H_1 E))</math></b>				
Zero out of four	-10.99	-9.64	-5.59	-4.24
One out of four	-6.59	-5.24	-3.89	-2.54
Two out of four	-2.2	-0.85	-2.2	-0.85
Three out of four	2.2	3.55	-0.5	0.85
Four out of four	6.59	7.94	1.19	2.54

### Model without individual differences

This probability weighting model assumes the absence of inter-individual differences with regard to the weighting parameters  $\gamma$  and  $p_0$ . The parameters  $\gamma$  and  $p_0$  were provided with weakly informative prior distributions—i.e., prior distributions restricting the posterior distribution very little—to ensure computational tractability (Gelman et al., 2014a).

The sampling statement for  $\gamma \sim N(1, 1) [0, \infty)$ , with  $N$  being the normal distribution truncated at zero. **Figure 3A** depicts the corresponding probability density. The prior distribution of  $\gamma$  is a normal distribution centered on 1, to represent the weak prior assumption of no probability weighting taking place. We chose a flat prior  $p_0 \sim \text{beta}(1, 1)$ , where *beta* is the beta function. **Figure 3B** shows the corresponding probability density.

Thus, all urn inferences are generated by the underlying Bernoulli distribution

$$y_n \sim \text{bern}(\text{logistic}(\gamma \cdot Lo(p) + (1 - \gamma) \cdot Lo(p_0))), \tag{12}$$

with  $p = P(H_1|E)$ , given a particular set of evidence  $E$ ,  $y_n \in \{0, 1\}$  with  $n \in [1, \dots, N]$ ,  $\gamma \in [0, \infty)$  and  $p_0 \in [0, 1]$ .  $\text{bern}(p_b)$  denotes the Bernoulli distribution:  $y_n$  takes on value one—denoting an inference of the urn with lower prior probability—with probability  $p_b$ , and  $y_n$  equals zero—an inference of the urn with higher prior probability—with probability  $1 - p_b$ .

**Figure 4** depicts the graphical model without individual differences. Nodes represent variables and arrows represent conditional probabilities. Discrete variables are given square nodes, continuous variables circular nodes and observed variables are shaded, while unobserved are not. Variables fully determined by their parents are given double borders, while stochastic variables are given single borders. Plate notation is used to ensure sparse representation, by grouping repeating variables in a subgraph enclosed by a rectangle, and indicating the number of repetitions. The plate corresponds to the  $N$  trials. An arrow between a variable outside of a plate to a variable inside indicates the dependence of each repeated variable on their parent outside the plate. The variables in the graphical model are defined using the subscript  $n$  to indicate definitions specific to the variables in each trial.

### Model with unrestricted individual differences

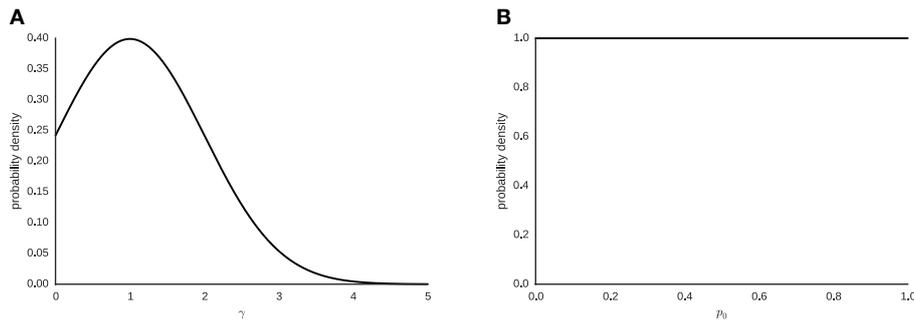
This model considers inter-individual variability in parameters  $\gamma_i$  and  $p_{0i}$ . The sampling statements of the prior distributions of each individual's parameters  $\gamma_i$  and  $p_{0i}$  are given by  $\gamma_i \sim N(1, 1) [0, \infty)$  and  $p_{0i} \sim \text{beta}(1, 1)$ .

The assumed data-generating process equals

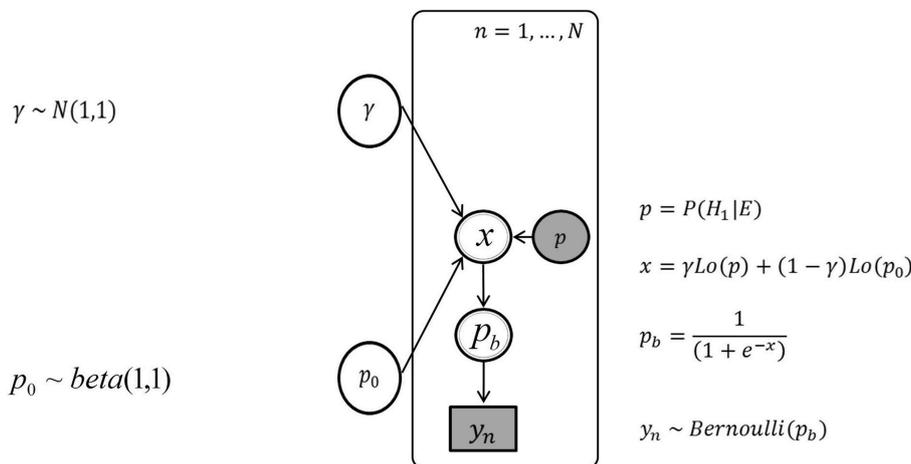
$$y_{ni} \sim \text{bern}(\text{logistic}(\gamma_i \cdot Lo(p) + (1 - \gamma_i) \cdot Lo(p_{0i}))), \tag{13}$$

with  $p = P(H_1|E)$ , given the corresponding set of evidence  $E$ ,  $y_{ni} \in \{0, 1\}$ ,  $\gamma_i \in [0, \infty)$ , and  $p_{0i} \in [0, 1]$  for each individual  $i$  and datum  $n$ .  $\text{bern}(p_b)$  denotes the Bernoulli distribution.

**Figure 5** shows the graphical model for the unrestricted inter-individual differences. The outer plate corresponds to the  $I = 16$  individuals and the inner plate to the  $N = 50$  trials. The variables



**FIGURE 3 | (A)** Gaussian prior distribution for parameter  $\gamma \sim N(1, 1)$  for the interval  $[0, \infty)$ . **(B)** Logit-normal prior distribution for parameter  $p_0 \sim \text{beta}(1, 1)$ .



**FIGURE 4 | The graph of the Bayesian network without inter-individual differences, with the urn inference  $y_n$ , the posterior probability  $p$ , and the parameters  $p_0$  and  $\gamma$ .**

in the graphical model are defined using the subscript  $i$  or  $n$  to indicate definitions for variables specific to each individual and trial, respectively.

### Model with hierarchical individual differences

In this modeling attempt, we assume that the latent  $\gamma$  and  $p_0$  parameters for individual participants are generated by more abstract latent parameters (hyper-parameters) describing group distributions across individuals. This model considers individual differences by specifying distributions of parameters  $\gamma$  and  $p_0$ , out of which individual  $\gamma_i$  and  $p_{0i}$  have to be sampled. For the parameter  $\gamma$ , these distributions were assumed to be normal distributions, characterized by mean  $\mu_\gamma$  and standard deviation  $\sigma_\gamma$ , with weakly informative prior distributions. The beta distribution of parameter  $p_0$  was specified by two hyper-parameters, one being a mean parameter  $\varphi = \alpha/(\alpha + \beta)$  and one being a total count parameter  $\lambda = \alpha + \beta$ , instead of the usual  $\alpha$  and  $\beta$ , following Gelman et al. (2014b).

The sampling statement for the parameter  $\gamma_i$  for each individual  $i$  was  $\gamma_i \sim N(\mu_\gamma, \sigma_\gamma) [0, \infty)$  with hyper-priors  $\mu_\gamma \sim N(1, 1)$  and  $\sigma_\gamma \sim \text{Uniform}(0, \infty)$ . The sampling

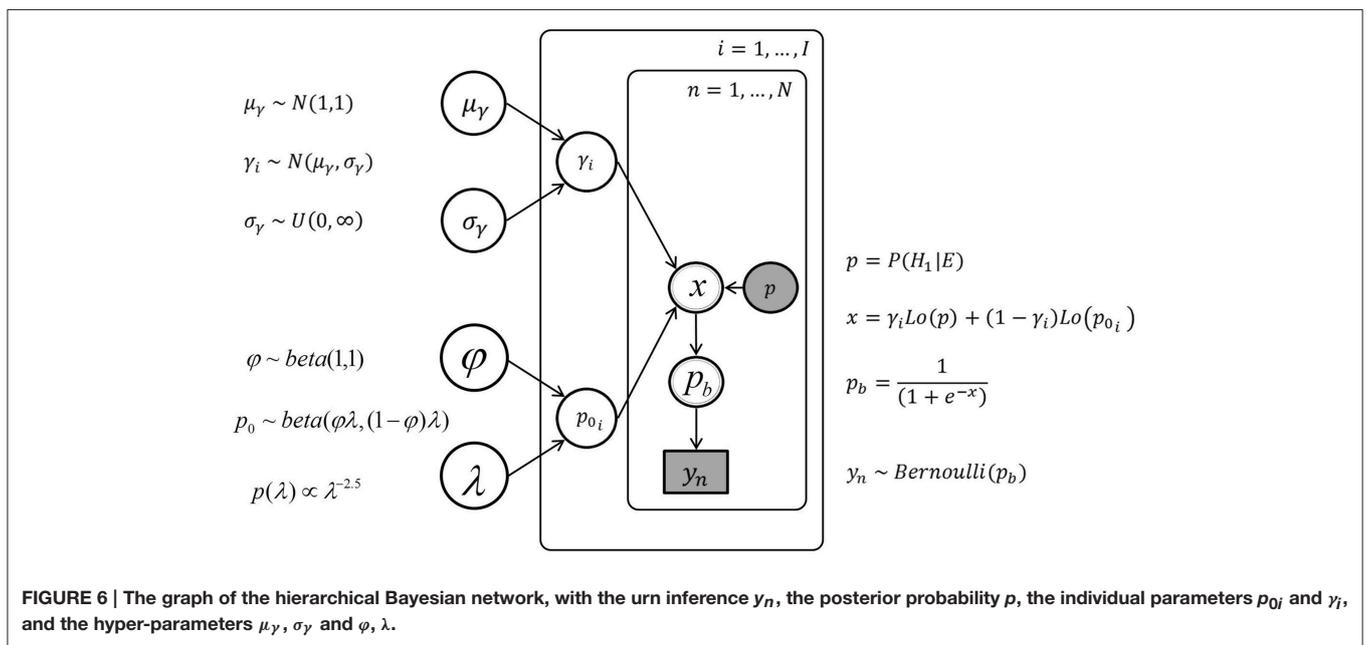
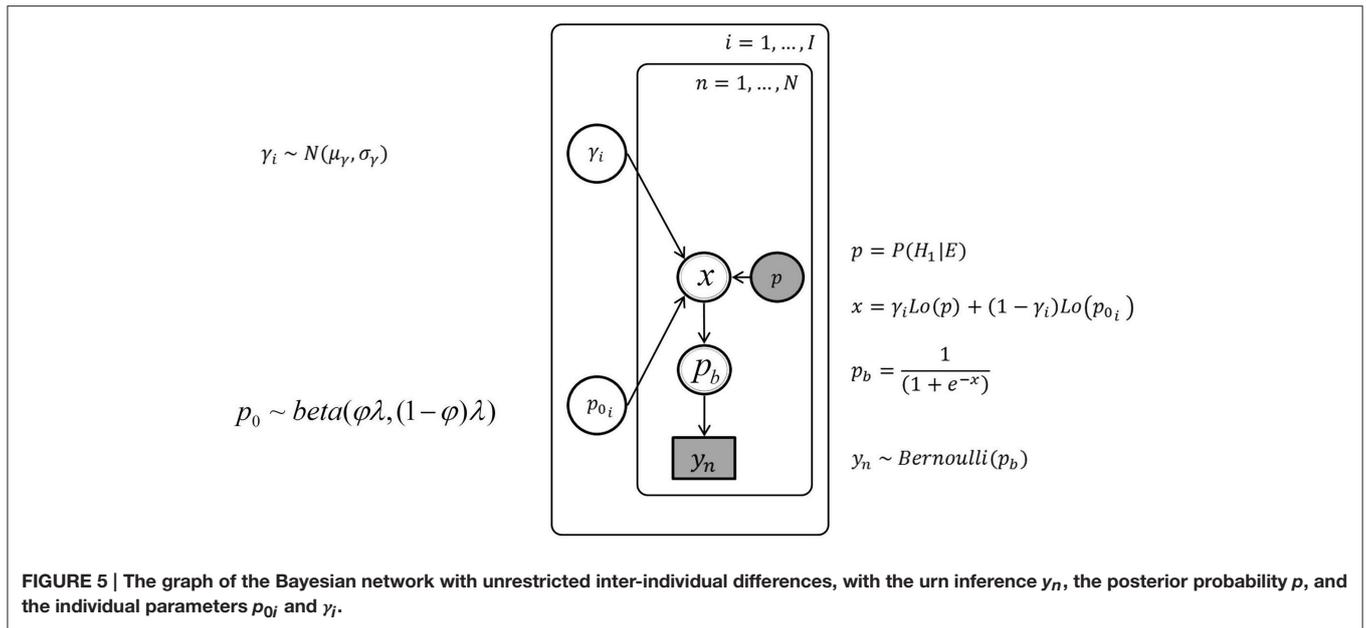
statement for the parameter  $p_{0i}$  for each individual  $i$  was  $p_{0i} \sim \text{beta}(\varphi\lambda, (1 - \varphi)\lambda)$  with  $\varphi \sim \text{beta}(1, 1)$  and  $\lambda$  having a pareto prior,  $p(\lambda) \propto \lambda^{-2.5}$ . The hierarchical model and the model with unrestricted individual differences share the same sampling state for the data (Equation 14), they differ only in the dependence (for the hierarchical model) or independence (for the model with unrestricted differences) of the  $\gamma_i$  and  $p_{0i}$  parameters. **Figure 6** displays the graphical model for the hierarchical individual differences.

### Model Evaluation and Selection

The quality of the out-of-sample prediction of the models was evaluated by comparing their (a) *widely applicable information criterion* (WAIC; Watanabe, 2010) and (b) log-likelihood in a leave-one out cross-validation. For both measures, the log point-wise predictive density (lppd) was computed as,

$$lppd = \sum_{n=1}^N \ln\left(\frac{1}{S} \cdot \sum_{s=1}^S P(y_n | \theta^s)\right) \quad (14)$$

Gelman et al. (2014b). This quantity measures the likelihood  $P(y_n | \theta^s)$  of the datum  $y_n$ , using  $s = 1, \dots, S$  sampling



results from the posterior distribution of all model parameters, labeled  $\theta^s$ .

**The Widely Applicable Information Criterion (WAIC)**

The WAIC uses the log point-wise predictive density and subtracts a correction term, corresponding to the effective number of parameters (Gelman et al., 2014b), here given by

$$\sum_{n=1}^N V_{s=1}^S(\ln P(y_n | \theta^s)), \tag{15}$$

with the likelihood  $P(y_n | \theta^s)$  of the datum  $y_n$ , using  $s = 1, \dots, S$  sampling results from the posterior distribution of all model parameters, labeled  $\theta^s$ , and the variance of the posterior,

$$V_{s=1}^S(a_s) = \frac{1}{S-1} \cdot \sum_{s=1}^S (a_s - \bar{a})^2. \tag{16}$$

Note that  $a_s$  is represented by  $\ln P(y_n | \theta^s)$  in (17).

Subtracting (15) from (14) gives

$$\sum_{n=1}^N \ln\left(\frac{1}{S} \cdot \sum_{s=1}^S P(y_n | \theta^s)\right) - \sum_{n=1}^N V_{s=1}^S(\ln P(y_n | \theta^s)), \tag{17}$$

which is the formula for calculating the WAIC (Gelman et al., 2014b). Multiplying the result from Equation (17) with  $-2$  gives us a measure on the scale of other deviance and information measures like deviance information criterion (DIC; Spiegelhalter et al., 2002) and an information criterion (AIC; Akaike, 1973).

### Bayesian Leave-One-Out Cross-Validation

In Bayesian leave-one-out cross-validation (Vehtari and Lampinen, 2002), each data point is predicted using the estimates of the posterior distributions obtained by fitting the model to all remaining data points. We fit the model with 15 individuals to predict the urn inferences of the 16-th person, since we are interested in predicting the urn inferences of additional individuals, rather than additional urn inferences of each individual. By doing this for every individual, we can calculate the sum of the individual log point-wise predictive densities. To make the interpretation of this score similar to other deviance measures, where smaller values denote a better fit, we use  $-2 \cdot lppd$ . The model assuming unrestricted individual differences makes no prediction about the individual parameters  $\gamma$  and  $p_0$  for additional individuals. It is therefore impossible to cross-validate this model, and this model fails a basic test of generalizability because it does not make sensible predictions for the behavior of additional individuals. For the model without individual differences, using the posterior distributions of parameters  $\gamma$  and  $p_0$  is straightforward to compute the log posterior predictive density given in Equation (14). For the model with hierarchical individual differences, the unknown individual parameters  $\gamma$  and  $p_0$  were assumed to be sampled out of the hierarchical distributions characterized by the hyper-parameters  $\mu_\gamma$ ,  $\sigma_\gamma$ ,  $\varphi$ , and  $\lambda$ .

The data were analyzed in Python 2.7.9 (Oliphant, 2007), and the models were fitted using the *Stan Hamiltonian-Monte-Carlo sampler* that provides approximate inference of the posterior distributions of the unknown parameters (Stan Development Team, 2013; Hoffman and Gelman, 2014). Each model in each condition was sampled with 10 chains and 20,000 iterations, with random initialization of the parameters and a warmup phase of 10,000 iterations (Data Sheet 2 in Supplementary Material).

## RESULTS

### Data Description

In **Figure 7**, the proportions of  $H_1$ -urn inferences are plotted separately for each condition ( $P_{cL_c}$ ,  $P_{uL_c}$ ,  $P_{cL_u}$ ,  $P_{uL_u}$ ) as a function of the log-odds posterior probabilities,  $Lo(P(H_1 | E))$ . Points represent the proportion from one participant; small random noise ( $\pm 0.25$ ) was added to the points to spread them out on the  $x$ -axis. Diamond symbols indicate mean proportions of inferences of  $H_1$  across participants, red dots indicate the proportion corresponding to the unweighted posterior probabilities. There is a considerable difference between the ideal proportion, indicated by the cross symbol, and the actual proportions of urn inferences of the participants, especially in the  $P_{cL_u}$  condition, and there is also considerable variation between the individual proportions. Not represented are the different

numbers of urn inferences which constituted the individual proportions.

### Parameter Estimation

#### Unweighted Bayesian Probabilities and Base Rate Neglect

Predictions of the unweighted Bayesian model depend solely on posterior probabilities,  $P(H_1|E)$ , whereas predictions of the base rate neglect model depend solely on the likelihood of the data,  $P(E|H_1)$ . The predictions of these two models therefore do not depend on additional parameters. Further information about the quality of these predictions is given below under *Model Evaluation and Selection*.

#### Weighted Bayesian Posterior Probabilities

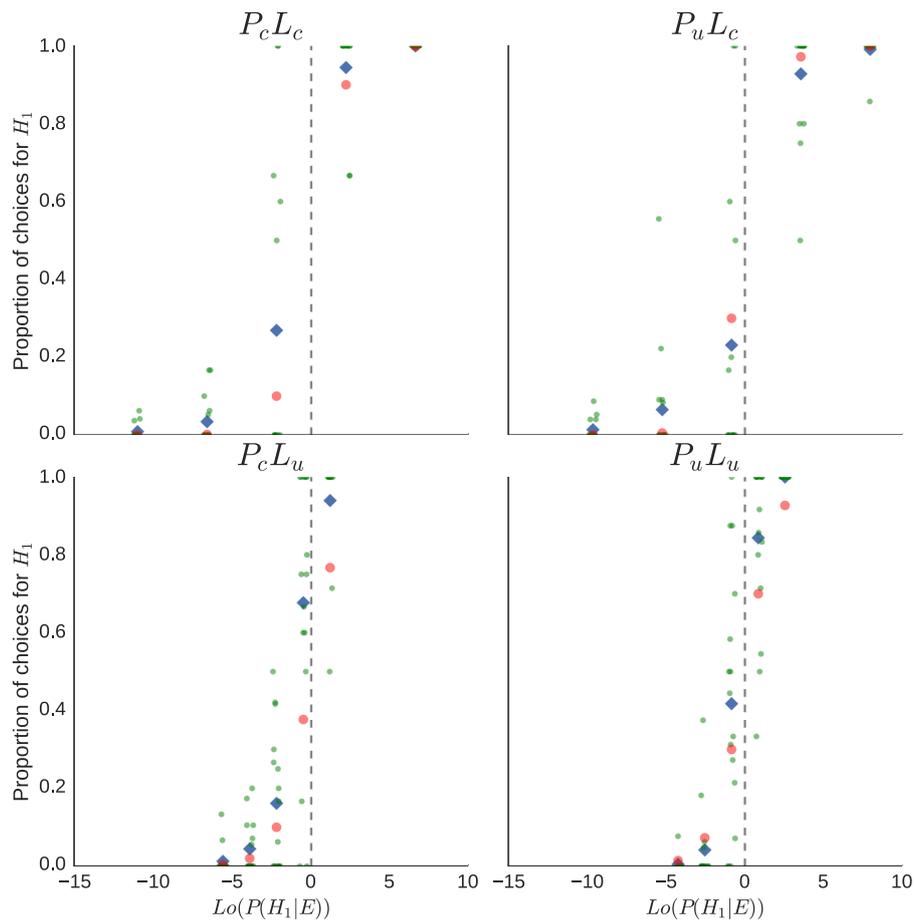
For the model without individual differences, sampling from the model described in Equation (12), yields the posterior distributions for parameters  $\gamma$  and  $p_0$  for each condition. The means of these posterior distributions were used as the estimates for the parameters in the weighting function described in Equation (5). The resulting functions vary across conditions, because the means of the posterior distributions for parameters  $\gamma$  and  $p_0$  differ between conditions. **Figure 8** displays the resulting probability weighting functions.

The estimated weighted probability functions differ between conditions, indicating different regions of over- or under-representation of probabilities, with the exception of condition  $P_{cL_u}$ , where, according to the mean of the posterior distribution of the model, very little probability weighting takes place. Descriptive statistics of the fitted model without inter-individual differences are attached (see Appendix A in Supplementary Material).

For the model with unrestricted individual differences, sampling from the model described in Equation (13), yields the posterior distributions of the parameters specific for each individual in each condition. In each condition, the means of the posterior distributions of the individual parameters  $\gamma$  and  $p_0$  were used as the estimates of the parameters of the individual weighting functions described in Equation (5). **Figure 9** shows the probability weighting functions for each individual and condition.

There is considerable variation in the shape of the probability weighting functions between individuals in one condition, and between conditions in general. Descriptive statistics of the fitted model with unrestricted inter-individual differences are attached (see Appendix B in Supplementary Material).

For the model with hierarchical individual differences, the posterior distributions of each individual parameter  $\gamma$  and  $p_0$ , and their hyper-parameters  $\varphi$ ,  $\lambda$ ,  $\mu_\gamma$ , and  $\sigma_\gamma$  for each condition were approximated by sampling. The means of the posterior distributions of  $\mu_\gamma$  and  $\sigma_\gamma$ , for each condition, were used to plot the expected hierarchical normal distribution of the individual parameters  $\gamma \sim N(\mu_\gamma, \sigma_\gamma)$  in **Figure 10**. The posterior distribution of the  $\gamma$  parameter varies considerably between conditions. For conditions with a certain likelihood the posterior is centered around one, and is quite narrow. This is in contrast to



**FIGURE 7 | Explorative data analysis for the individual proportions of  $H_1$ -inferences per participant and sequence of drawn balls.** Each point represents the proportion of an individual per sequence, diamond shapes indicate the mean proportion of  $H_1$ -inferences per sequence, and red dots indicate the appropriate proportion of  $H_1$ -inferences, given the posterior probability  $P(H_1|E)$ . Overlapping proportions are “jittered” by adding small random noise on the x-axis.

the conditions with an uncertain likelihood, where the posterior is centered around higher values of  $\gamma$  and more spread out.

The means of the posterior distributions of  $\varphi$  and  $\lambda$ , for each condition, were used to plot the expected hierarchical beta distribution of the individual parameters  $p_0 \sim \text{beta}(\varphi\lambda, (1 - \varphi)\lambda)$  (Figure 11). We can distinguish two patterns in Figure 11: The posterior mode is (slightly) more extreme for certain than for uncertain prior conditions, and the posterior mode is closer to zero in conditions with an uncertain likelihood and closer to one in conditions with a certain likelihood.

For the individual parameters  $\gamma$  and  $p_0$  for each individual and condition, the means of the posterior distributions were used as the parameters of the weighting function described in Equation (5), and the resulting functions were plotted in Figure 12.

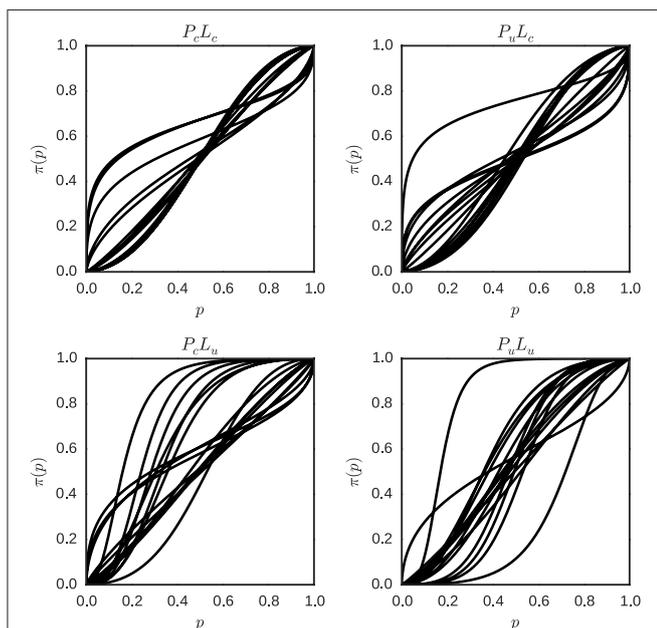
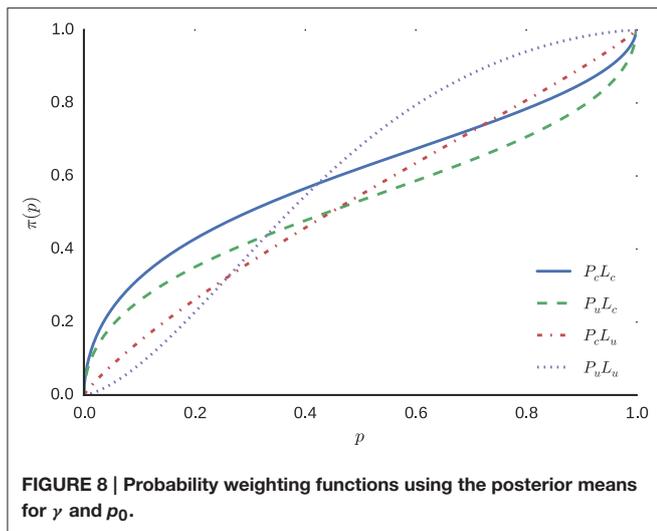
There is still variation in the shape of the probability weighting functions between conditions, but the variation between individuals in one condition is reduced in contrast to Figure 9 for conditions with a more certain likelihood. Interestingly, in conditions with a more uncertain likelihood the variation between participants is either similar ( $P_c L_u$ ) or larger

( $P_u L_u$ ). Descriptive statistics of the fitted model with hierarchical inter-individual differences are attached (see Appendix C in Supplementary Material).

## Model Evaluation and Selection

### Model Comparison Using the WAIC

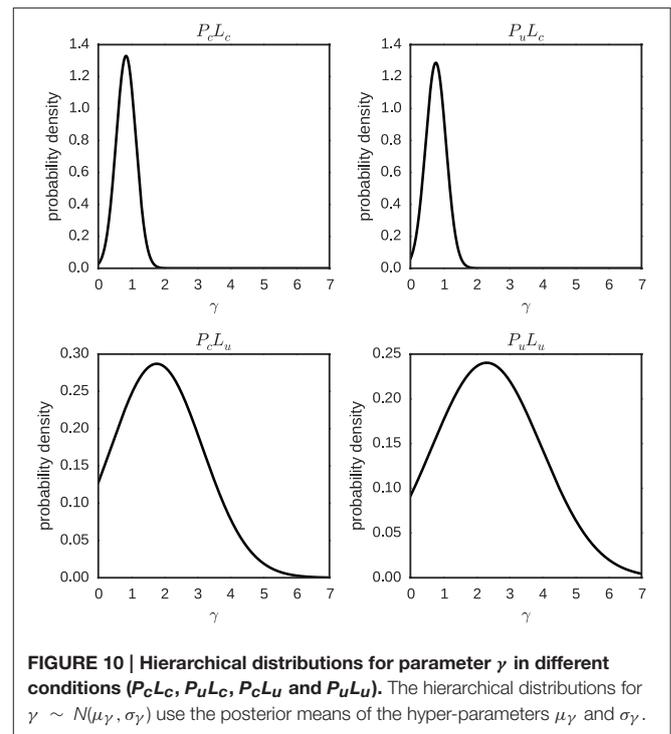
Table 2 displays the WAICs for different models for each condition ( $P_c L_c$ ,  $P_u L_c$ ,  $P_c L_u$ ,  $P_u L_u$ ). Smaller values indicate a better out-of-sample prediction than larger values. A comparison between models reveals superior performance of weighted Bayesian models compared to unweighted models. In the latter category, both models have similar WAIC values on uncertain prior conditions  $P_u$ , while the base rate neglect model performs better on the  $P_c L_c$  condition, and the unweighted Bayesian model outperforms the base rate neglect model on the  $P_c L_u$  condition. Among the weighted Bayesian models, those models with unrestricted individual differences and hierarchical individual differences perform better than the model without individual differences on all conditions. However, the hierarchical model outperforms the model with unrestricted



inter-individual differences on  $P_cL_u$  and  $P_uL_u$  conditions, while both models perform similarly on the remaining conditions. In sum, the unweighted models yield larger WAIC values than the weighted models, with the hierarchical model providing the best overall performance.

### Model Comparison Using Cross-Validation

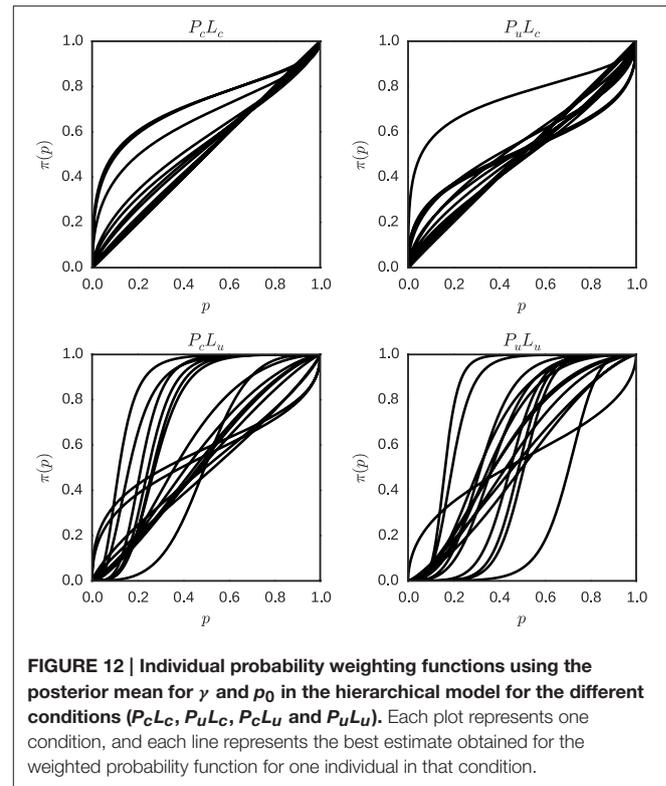
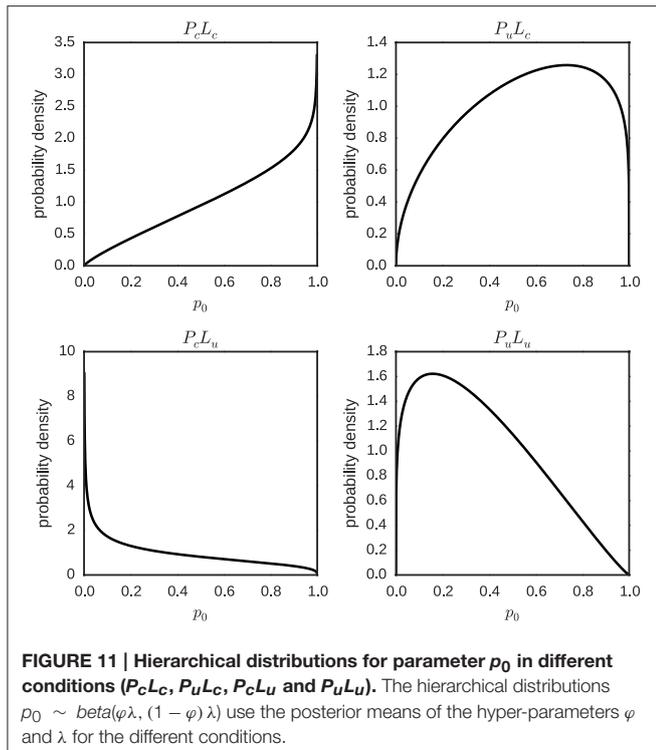
The results of the cross-validation, averaged over the 16 participants, are given in **Table 3** for each condition ( $P_cL_c$ ,  $P_uL_c$ ,  $P_cL_u$ ,  $P_uL_u$ ). Further, cross-validation was not applicable



to the model with unrestricted individual differences. The weighted Bayesian models predict new data better than the unweighted models. However, in the  $P_uL_u$  condition, the model without inter-individual differences yields the worst average prediction of all models, while the model with hierarchical inter-individual differences still yields the best prediction. Among the weighted Bayesian models, the model with hierarchical individual differences outperforms the model without differences on conditions  $P_uL_c$ ,  $P_cL_u$ , and  $P_uL_u$ , with especially large differences in the quality of prediction in uncertain likelihood conditions ( $L_u$ ), and similar performance of both models in the  $P_cL_c$  condition.

### Effects of Prior and Likelihood Manipulations

Since we observe differences in the probability weighting functions between conditions, we can test for the effect of the experimental prior and likelihood manipulations on the model parameters parameter  $\gamma$  and  $\rho_0$ . We use the parameters of the model with full inter-individual differences, and estimate a linear mixed-effects model, with a random intercept per participant, categorical variables representing the manipulation of prior and likelihood and the interaction between these manipulations (see Appendix D in Supplementary Material for a more detailed description and the full model statistics). We estimate such a model separately for  $\gamma$  and  $\rho_0$  and find a significant effect for the uncertain likelihood conditions in  $\gamma$  ( $\beta = 0.35$ ,  $SE = 0.17$ ,  $z = 2.01$ ,  $p = 0.036$ ). Uncertain likelihood conditions increase the  $\gamma$  coefficient and change the probability weighting function to an S-shape (or a convex shape if  $\rho_0$  is close to one). For the  $\rho_0$  parameter, we find a significant effect also only for uncertain



likelihood conditions ( $\beta = -0.15$ ,  $SE = 0.05$ ,  $z = -3.00$ ,  $p = 0.003$ ). In uncertain likelihood conditions the cutting point of the probability weighting function is lower than in certain likelihood conditions. Of course, this effect can also be seen in the posterior of the hierarchical model.

## DISCUSSION

In this article, we explored various possibilities for modeling cognitive processes for probabilistic inference. To that end, probabilistic inference was observed in a small world (Savage, 1954), vested as a classic urn-ball paradigm (Phillips and Edwards, 1966; Grether, 1980, 1992; Stern et al., 2010; Achtziger et al., 2014) involving a factorial two (prior probabilities) by two (likelihoods) design. Probabilistic inference was modeled as originating from different variants of Bayesian inference. Five computational models of cognitive processes for probabilistic inference were compared by Bayesian model evaluation (Vehtari and Lampinen, 2002; Shiffrin et al., 2008). We found considerable task dependency such that more certain likelihoods were associated with mean group probability weighting parameters that satisfy  $\gamma < 0.6$  (cf. Tables A.1, A.2), whereas less certain likelihoods were associated with mean group probability weighting parameters that satisfy  $\gamma < 0.9$  (cf. Table A.3) or  $\gamma > 1$  (cf. Table A.4). We are not aware of a psychological theory that could predict this strong task dependency of probability weighting.

The least complex models (i.e., parameter-free Bayesian posterior probabilities, parameter-free base rate neglect) provide inadequate models of probabilistic inference in terms of the

robustness of the fit between models and data as well as in terms of their generalizability. The introduction of probability weighting functions (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992; Prelec, 1998; Gonzalez and Wu, 1999; Luce, 2000; Zhang and Maloney, 2012; Cavagnaro et al., 2013) yielded more robust and generalizable fits between models and data than the two parameter-free models. Probability weighting models share the assumption that subjective probabilities deviate from true probabilities due to (inverted) S-shaped distortions of probabilities.

The least complex of these models conceptualized different slope,  $\gamma$ , and crossover point,  $p_0$ , parameters for probability distortion across conditions, but not across individuals (model without individual differences). This model includes eight free parameters (four conditions by two free parameters, i.e., four  $\gamma$  and four  $p_0$  parameters).

There were two variants of individual difference models which envisaged different slope and crossover point parameters for probability distortion across conditions and individuals. The unrestricted individual differences model considered individual differences, by specifying individual  $\gamma_i$  and  $p_{0,i}$  parameters, which are fully determined by the data. This model includes four (conditions) by 32 free parameters (16 individual  $\gamma_i$  parameters and 16 individual  $p_{0,i}$  parameters, for a total of 128 free parameters).

In contrast, the hierarchical individual differences model assumed individual parameters themselves to be generated by more abstract latent parameters (hyper-parameters) describing group distributions across individuals. This model thus

**TABLE 2 | WAIC deviance measures and standard errors for experimental conditions and models.**

Model	Condition			
	$P_{cLc}$	$P_{uLc}$	$P_{cLu}$	$P_{uLu}$
<b>Unweighted Bayesian</b>				
WAIC	157.90	207.70	257.38	293.58
SE	60.30	64.34	37.09	37.77
<b>Base rate neglect</b>				
WAIC	123.28	203.73	356.66	294.37
SE	32.33	54.62	36.00	28.67
<b>WEIGHTED BAYESIAN MODELS</b>				
<b>No differences</b>				
WAIC	93.90	139.93	228.48	263.24
SE	15.50	23.37	24.52	46.66
<b>Unrestricted differences</b>				
WAIC	71.59	118.03	162.20	187.89
SE	13.60	18.01	17.04	22.34
<b>Hierarchical differences</b>				
WAIC	73.08	121.60	152.77	163.08
SE	13.19	18.17	16.37	22.86

considered individual differences by specifying distributions of parameters  $\gamma$  and  $p_0$ , out of which individual  $\gamma_i$  and  $p_{0i}$  had to be sampled. This model includes four (conditions) by 32 free parameters (16 individual  $\gamma_i$  parameters and 16 individual  $p_{0i}$  parameters), plus sixteen hyper-parameters (four conditions by four hyper-parameters,  $\mu_\gamma$ ,  $\sigma_\gamma$ ,  $\mu_{p_0}$ ,  $\sigma_{p_0}$ , for a total of 144 parameters). Since the effective number of parameters in a hierarchical model depends on the variance of the group level parameters (Gelman et al., 2014b), the actual number of free parameters is  $<144$ . We show that the hierarchical individual differences model outperformed the unrestricted individual differences model, which in turn outperformed the model without individual differences. This is the result of a process of model building, in which each increase in model complexity is justified by an increase in the model's fit and predictive power. To conclude, the assumption of large differences in probability distortions across tasks and individuals (i.e., in the values of the slope and crossover point parameters) seem critical for understanding Bayesian inference (see also Glöckner and Pachur, 2012; Zhang and Maloney, 2012). Further, it seems advantageous to consider individual differences as being sampled from weakly informative prior distributions of individual parameter values.

The hierarchical model of weighted Bayesian posterior probabilities is thus the preferable model of probabilistic inference, despite the non-negligible association between model performance and model complexity. However, parameter counts and estimations of model complexity should not be considered as equivalents. While the hierarchical model is more complex in terms of the number of parameters, it also restricts the individual parameters more than the model with unrestricted differences. Further, parameters, which are fully determined by the data

**TABLE 3 | Log posterior predictive density of leave-one-out cross-validation and standard errors for experimental conditions and models.**

Model	Condition			
	$P_{cLc}$	$P_{uLc}$	$P_{cLu}$	$P_{uLu}$
<b>Unweighted Bayesian</b>				
LOO-CV	157.92	207.68	257.44	293.60
SE	60.30	64.34	37.09	37.77
<b>Base rate neglect</b>				
LOO-CV	123.2	203.68	356.64	294.4
SE	32.33	54.61	36.00	28.66
<b>WEIGHTED BAYESIAN MODELS</b>				
<b>No differences</b>				
LOO-CV	89.28	133.12	215.36	248.64
SE	16.89	23.58	15.66	30.57
<b>Unrestricted differences</b>				
	n.a.	n.a.	n.a.	n.a.
<b>Hierarchical differences</b>				
LOO-CV	95.36	126.08	200.32	182.4
SE	22.74	20.66	27.02	15.18

(or by parts of the data, as in the unrestricted model), have a higher chance to overfit in comparison to those parameters that are imposed by some theoretical structure (e.g., the similarity of inter-individual differences, as in the hierarchical model). Similarly, to further prevent overfitting, the hierarchical model could be extended to explicitly model the different experimental conditions through an additional set of hyper-parameters. More concretely, an implementation in the framework of hierarchical Bayesian models would model the influence of subject, as well as experimental manipulation on the parameters of a probability weighting function. Further potential confounds, like task-order, and sequential effects in probability weighting might also be included. A final expansion of the hierarchical model could be its formulation as a Bayesian hierarchical mixture model (Bartlema et al., 2014), allowing groups of participants to have discrete, in addition to continuous, inter-individual differences.

Hierarchical Bayesian modeling offers specific advantages. Since knowledge about parameters propagates in hierarchical Bayesian models, the flow of probabilistic influence (Koller and Friedman, 2009) allows individual parameters to influence each other with regard to the certainty of their estimation. This renders the sharing of statistical strength (Gelman et al., 2014a) possible, with consequential improvements in predictive accuracy under high levels of uncertainty. This ubiquitous characteristic of hierarchical Bayesian modeling can also be recognized in our study, where the hierarchical model of weighted Bayesian posterior probabilities proved especially superior on those conditions that involved particularly high uncertainty about individual parameter values (i.e., the  $L_u$  conditions).

Our results are highly dependent on task conditions and individuals. However, the results from hierarchical Bayesian modeling converge with previous results in revealing that

probability weighting parameters show considerable task dependency and individual differences (Tversky and Kahneman, 1992; Gonzalez and Wu, 1999; Stott, 2006; Wu et al., 2009). We have to admit that our work does not offer an a-priori explanation nor a post-hoc speculation as to why this should be the case. The hierarchical model is hence not a theory of cognitive processes for probabilistic inference, despite its hereby documented success in providing a reasonably good estimate of the observed data. Such a theory would encompass an answer to the question why subjective probabilities may deviate from the true probabilities in the ways that they apparently do, and specify the factors that affect this probability distortion. A full explanation of the phenomena just described would require not only that we account for the S-shaped form of the probability distortion, but also for the large differences in the values of the slope and crossover point parameters across tasks and individuals. Although such a theory still needs to be developed, we introduced hierarchical Bayesian modeling as a valuable method for developing a model, for testing it against data, for checking if there is systematic error in the predictions of the model, for increasing the complexity of the model as long as there are intolerable amounts of error, and for checking whether the more complex model provides a superior explanation for the observed data. Furthermore, hierarchical Bayesian modeling of cognitive processes for probabilistic inference may serve an instrumental role in neuroscientific studies of Bayesian inference (Kolossa et al., 2013, 2015).

## CONCLUSION

Our findings also bear on the question whether inductive inference (Anderson, 2015) can be described as being Bayesian or not (Gigerenzer and Murray, 1987; Koehler and Harvey, 2004). According to Clark (2013), the answer to this question is empirically underdetermined; thus, new methods for tackling the problem need to be developed and evaluated. Historically, two almost certainly irreconcilable intellectual camps evolved, although serious attempts exist to bridge the conflictive positions

## REFERENCES

- Achtziger, A., Alós-Ferrer, C., Hügelschäfer, S., and Steinhauser, M. (2014). The neural basis of belief updating and rational decision making. *Soc. Cogn. Affect. Neurosci.* 9, 55–62. doi: 10.1093/scan/nss099
- Akaike, H. (1973). "Information theory and an extension of the maximum likelihood principle," in *Proceeding of the Second International Symposium on Information Theory*, eds B. N. Petrov and F. Caski (Budapest: Akademiai Kiado), 267–281.
- Anderson, J. R. (2015). *Cognitive Psychology and Its Implications, 8th Edn.* London, UK: Palgrave Macmillan.
- Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychol.* 44, 211–233. doi: 10.1016/0001-6918(80)90046-3
- Bartlema, A., Lee, M. D., Wetzels, R., and Vanpaemel, W. (2014). A Bayesian hierarchical mixture approach to individual differences: case studies in selective attention and representation in category learning. *J. Math. Psychol.* 59, 132–150. doi: 10.1016/j.jmp.2013.12.002
- Cavagnaro, D. R., Pitt, M. A., Gonzalez, R., and Myung, J. I. (2013). Discriminating among probability weighting functions using adaptive design optimization. *J. Risk Uncertain.* 47, 255–289. doi: 10.1007/s11166-013-9179-3
- Chater, N., Oaksford, M., Hahn, U., and Heit, E. (2010). Bayesian models of cognition. *Wiley Interdiscip. Rev. Cogn. Sci.* 1, 811–823. doi: 10.1002/wcs.79
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477
- Edwards, W. (1954). The theory of decision making. *Psychol. Bull.* 51, 380–417. doi: 10.1037/h0053870
- Edwards, W. (1982). "Conservatism in human information processing," in *Judgment Under Uncertainty: Heuristics and Biases*, eds D. Kahneman, P. Slovic and A. Tversky (Cambridge: Cambridge University Press), 359–369.
- Edwards, W., Lindman, H., and Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychol. Rev.* 70, 193–242. doi: 10.1037/h0044139
- Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.* 59, 255–278. doi: 10.1146/annurev.psych.59.103006.093629
- Evans, J. S. B., and Stanovich, K. E. (2013). Dual-process theories of higher cognition advancing the debate. *Perspect. Psychol. Sci.* 8, 223–241. doi: 10.1177/1745691612460685
- Gallistel, C. R. (2009). The importance of proving the null. *Psychol. Rev.* 116, 439–453. doi: 10.1037/a0015251

in the form of dual process models (Evans, 2008; Evans and Stanovich, 2013). The Bayesian camp claimed that there is sufficient evidence for postulating that Bayes' theorem can serve as a descriptive theory of human inductive inference (Chater et al., 2010; Tenenbaum et al., 2011; Griffiths et al., 2012). However, the non-Bayesian camp insisted that the evidence showed that humans critically deviate from prescriptive Bayesian solutions (e.g., base rate neglect; Kahneman and Tversky, 1973; Bar-Hillel, 1980). In fact, some researchers believe that humans should be better described as *homo heuristicus* rather than as rational probabilists (Kahneman et al., 1982; Gigerenzer and Brighton, 2009). Our data offer another possibility: Humans might be able to integrate information for inductive inference according to Bayesian prescriptions (cf. Equations 1–4), yet in terms of distorted probabilities (cf. Equations 5–9), as originally conjectured by Edwards (1954) and by Prospect Theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992).

## AUTHOR CONTRIBUTIONS

BK developed the study design and collected the data. MB performed the data analysis under the supervision of BK. MB and BK drafted the manuscript, and CS and FL provided critical revisions. All authors approved the final version of the manuscript for submission.

## FUNDING

The research reported here was supported by a grant to BK from the Petermax-Müller-Stiftung, Hannover, Germany. In addition, FL received funding from the German National Academic Foundation.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00755>

- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014a). *Bayesian Data Analysis, Vol. 2*. London: Chapman & Hall/CRC.
- Gelman, A., Hwang, J., and Vehtari, A. (2014b). Understanding predictive information criteria for Bayesian models. *Stat. Comput.* 24, 997–1016. doi: 10.1007/s11222-013-9416-2
- Gelman, A., and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Stat. Sci.* 7, 457–472.
- Gigerenzer, G., and Brighton, H. (2009). Homo heuristicus: why biased minds make better inferences. *Top. Cogn. Sci.* 1, 107–143. doi: 10.1111/j.1756-8765.2008.01006.x
- Gigerenzer, G., and Murray, D. J. (1987). *Cognition as Intuitive Statistics*. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.
- Glöckner, A., and Pachur, T. (2012). Cognitive models of risky choice: parameter stability and predictive accuracy of prospect theory. *Cognition* 123, 21–32. doi: 10.1016/j.cognition.2011.12.002
- Gold, J. I., and Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.* 30, 535–574. doi: 10.1146/annurev.neuro.29.051605.113038
- Gonzalez, R., and Wu, G. (1999). On the shape of the probability weighting function. *Cogn. Psychol.* 38, 129–166. doi: 10.1006/cogp.1998.0710
- Grether, D. M. (1980). Bayes rule as a descriptive model: the representativeness heuristic. *Q. J. Econ.* 95, 537–557. doi: 10.2307/1885092
- Grether, D. M. (1992). Testing Bayes rule and the representativeness heuristic: some experimental evidence. *J. Econ. Behav. Organ.* 17, 31–57. doi: 10.1016/0167-2681(92)90078-P
- Grether, D. M., and Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *Am. Econ. Rev.* 69, 623–638.
- Griffiths, T. L., Vul, E., and Sanborn, A. N. (2012). Bridging levels of analysis for probabilistic models of cognition. *Curr. Dir. Psychol. Sci.* 21, 263–268. doi: 10.1177/0963721412447619
- Hoffman, M. D., and Gelman, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* 15, 1593–1623.
- Hojitink, H. (2012). *Informative Hypotheses: Theory and Practice for Behavioral and Social Scientists*. Boca Raton, FL: CRC Press.
- Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press.
- Johnson, J. G., and Busemeyer, J. R. (2010). Decision making under risk and uncertainty. *Wiley Interdiscip. Rev. Cogn. Sci.* 1, 736–749. doi: 10.1002/wcs.76
- Kahneman, D., Slovic, P., and Tversky, A. (eds.) (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.
- Kahneman, D., and Tversky, A. (1973). On the psychology of prediction. *Psychol. Rev.* 80, 237–251. doi: 10.1037/h0034747
- Kahneman, D., and Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–291. doi: 10.2307/1914185
- Knight, F. H. (1921). *Risk, Uncertainty and Profit*. New York, NY: Hart, Schaffner and Marx.
- Koehler, D. J., and Harvey, N. (2004). *Blackwell Handbook of Judgment and Decision Making*. Oxford: Blackwell Publishing.
- Koehler, J. J. (1996). The base rate fallacy reconsidered: descriptive, normative, and methodological challenges. *Behav. Brain Sci.* 19, 1–17. doi: 10.1017/S0140525X00041157
- Koller, D., and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA: MIT Press.
- Kolossa, A., Fingscheidt, T., Wessel, K., and Kopp, B. (2013). A model-based approach to trial-by-trial P300 amplitude fluctuations. *Front. Hum. Neurosci.* 6:359. doi: 10.3389/fnhum.2012.00359
- Kolossa, A., Kopp, B., and Fingscheidt, T. (2015). A computational analysis of the neural bases of Bayesian inference. *Neuroimage* 106, 222–237. doi: 10.1016/j.neuroimage.2014.11.007
- Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends Cogn. Sci.* 14, 293–300. doi: 10.1016/j.tics.2010.05.00
- Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *J. Math. Psychol.* 55, 1–7. doi: 10.1016/j.jmp.2010.08.013
- Luce, R. D. (2000). *Utility of Gains and Losses: Measurement-Theoretical and Experimental Approaches*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- MacKay, D. J. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press.
- Oliphant, T. E. (2007). Python for scientific computing. *Comput. Sci. Eng.* 9, 10–20. doi: 10.1109/MCSE.2007.58
- Phillips, L. D., and Edwards, W. (1966). Conservatism in a simple probability inference task. *J. Exp. Psychol.* 72, 346–354. doi: 10.1037/h0023653
- Prelec, D. (1998). The probability weighting function. *Econometrica* 66, 497–527. doi: 10.2307/2998573
- Robert, C. (2007). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation, 2nd Edn*. New York, NY: Springer.
- Savage, L. J. (1954). *The Foundations of Statistics*. New York, NY: Wiley.
- Seer, C., Lange, F., Boos, M., Dengler, R., and Kopp, B. (2016). Prior probabilities modulate cortical surprise responses: a study of event-related potentials. *Brain Cogn.* doi: 10.1016/j.bandc.2016.04.011
- Shanteau, J. (1972). Descriptive versus normative models of sequential inference judgment. *J. Exp. Psychol.* 93, 63–68. doi: 10.1037/h0032509
- Shanteau, J. (1975). Averaging versus multiplying combination rules of inference judgment. *Acta Psychol.* 39, 83–39. doi: 10.1016/0001-6918(75)90023-2
- Shiffrin, R. M., Lee, M. D., Kim, W., and Wagenmakers, E. J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cogn. Sci.* 32, 1248–1284. doi: 10.1080/03640210802414826
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *J. Roy. Stat. Soc. B* 64, 583–639. doi: 10.1111/1467-9868.00353
- Stan Development Team (2013). *Stan: A C++ Library for Probability and Sampling, Version 1.3*. Available online at: <http://mc-stan.org/>
- Stern, E. R., Gonzalez, R., Welsh, R. C., and Taylor, S. F. (2010). Updating beliefs for a decision: neural correlates of uncertainty and underconfidence. *J. Neurosci.* 30, 8032–8041. doi: 10.1523/JNEUROSCI.4729-09.2010
- Stott, H. (2006). Prospect theory's functional menagerie. *J. Risk Uncertain.* 32, 101–130. doi: 10.1007/s11166-006-8289-6
- Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to grow a mind: statistics, structure, and abstraction. *Science* 331, 1279–1285. doi: 10.1126/science.1192788
- Tversky, A., and Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *J. Risk Uncertain.* 5, 297–323. doi: 10.1007/BF00122574
- Vehtari, A., and Lampinen, J. (2002). Bayesian model assessment and comparison using cross-validation predictive densities. *Neural Comput.* 14, 2439–2468. doi: 10.1162/08997660260293292
- Volz, K. G., and Gigerenzer, G. (2012). Cognitive processes in decisions under risk are not the same as in decisions under uncertainty. *Front. Neurosci.* 6:105. doi: 10.3389/fnins.2012.00105
- Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of *p* values. *Psychon. Bull. Rev.* 14, 779–804. doi: 10.3758/BF03194105
- Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* 11, 3571–3594.
- Wu, S. W., Delgado, M. R., and Maloney, L. T. (2009). Economic decision-making compared with an equivalent motor task. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6088–6093. doi: 10.1073/pnas.0900102106
- Zhang, H., and Maloney, L. T. (2012). Ubiquitous log odds: a common representation of probability and frequency distortion in perception, action, and cognition. *Front. Neurosci.* 6:1. doi: 10.3389/fnins.2012.00001

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Boos, Seer, Lange and Kopp. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.