Bayesian Methods Applied to the Generalized Matching Law

Manuel Villarreal¹, Carlos Velázquez¹, José L. Baroja¹, Alejandro Segura³, Arturo Bouzas¹, and Michael D. Lee²

¹Facultad de Psicología, Universidad Nacional Autónoma de México
²Department of Cognitive Sciences, University of California Irvine
³Universidad de Guadalajara, Centro de Estudios e Investigaciones en Comportamiento

Abstract

We demonstrate the usefulness of Bayesian methods in developing, evaluating, and using psychological models in the experimental analysis of behavior. We do this through a case study, involving new experimental data that measure the response count and time allocation behavior in pigeons under concurrent random-ratio random-interval schedules of reinforcement. To analyze these data, we implement a series of behavioral models, based on the generalized matching law, as graphical models, and use computational methods to perform fully Bayesian inference. We demonstrate how Bayesian methods, implemented in this way, make inferences about parameters representing psychological variables, how they test the descriptive adequacy of models as accounts of behavior, and how they compare multiple competing models. We also demonstrate how the Bayesian graphical modeling approach allows for more complicated modeling structures, including hierarchical, common cause, and latent mixture structures, to formalize more complicated behavioral models. As part of the case study, we demonstrate how the statistical properties of Bayesian methods allow them to provide more direct and intuitive tests of theories and hypotheses, and how they support the creative and exploratory development of new theories and models.

Keywords: Bayesian methods, graphical models, generalized matching law, concurrent schedules
Introduction

A review of the publications in this journal conducted by Mazur (2006) noted the increasing use and interest, beginning in the 1960s, in mathematical modeling of the experimental analysis of behavior. These mathematical models have often been analyzed using simple methods, without a focus on statistical rigor. Typically, for example, parameter estimates are found by least-squares optimization (e.g., Aparicio, Baum, Hughes, & Pitts, 2016; Baum & Davison, 2014; Davison & McCarthy, 1988), and sometimes hypotheses are simply tested by observing that parameter estimates are close (or far) from a theoretically-expected value (e.g., Baum, 1979, 1983; Mullins, Agunwamba, & Donohoe, 1982).

In the last twenty-or-so years, however, many areas of psychology involving mathematical and statistical models have moved toward the widespread adoption of Bayesian methods (Kruschke, 2011; Lee, 2018). There are now many examples of psychological models—in areas including perception, memory, learning, development, categorization, decision making, and problem solving—being implemented and evaluated within the Bayesian framework. The textbooks by Farrell and Lewandowsky (2018) and Lee and Wagenmakers (2013) present a number of case studies spanning many of these domains, and edited volumes summarizing the field, such as Wixted and Wagenmakers (2018), emphasize the increasing role of Bayesian methods. In addition, the journal *Psychonomic Bulletin & Review* recently published an extensive special issue entirely devoted to Bayesian methods (Vandekerckhove, Rouder, & Kruschke, 2018).

Advantages of Bayesian methods

One of the major reasons for the recent increased use of Bayesian methods is technological. The development of computational methods for implementing Bayesian analyses has made them practical to use. Another major reason for psychological modeling adopting Bayesian methods is theoretical, involving a recognition of the statistical advantages and capabilities they offer (Jaynes, 2003; Lindley, 1972). The key principle of the Bayesian approach is that uncertainty, whether it is about latent psychological variables or predictions about behavior, is always represented by probability distributions. As more relevant information becomes available, usually in the form of additional data, these distributions are updated. The updating means that Bayesian methods always represent what is known and unknown about parameters, and are always capable of making probabilistic predictions about data based on what is known. In this way, Bayesian methods mirror the conceptual process of empirical science, as one of making progress by testing the predictions of models against empirical evidence (Feynman, 1994).

Perhaps the most compelling reason for the increased use of Bayesian methods, however, is the recognition of the new theoretical and modeling possibilities they afford. Bayesian in-
ference provides a simple and general framework for relating any probabilistic psychological model to data. This generality means that more theoretically complicated and phenomenologically complete psychological models can be considered. In particular, there has been a significant growth in hierarchical models, also called multi-level models, which provide more detailed accounts of the heterogeneity in behavior, and remain naturally amenable to Bayesian inference (Farrell & Lewandowsky, 2018; Lee & Wagenmakers, 2013; Young, 2018). In addition, because Bayesian methods of model evaluation account for both descriptive adequacy and complexity, there is a safe-guard against over-fitting the available empirical evidence. Together, the allowance of theoretical freedom and the requirement of empirical justification provide an ideal framework for attempting a more ambitious model-based understanding of human and animal behavior.

Overview of the case study

Against this background, the goal of this article is to demonstrate the advantages of Bayesian modeling in behavioral research, with a focus on the sort of research done in the experimental analysis of behavior. We do this by presenting a case study involving the generalized matching law. We first describe the law and emphasize its interpretation as a generative psychological model, making assumptions about how the psychological variables of sensitivity and bias control behavior. We then describe the data we use in the case study, which are new, and come from a behavioral experiment involving six pigeons on different pairs of concurrent random-ratio (RR) random-interval (RI) schedules.

The case study consists of three modeling analyses that apply the generalized matching law to the pigeon data. In the first modeling analysis, we show how Bayesian methods can be used to implement standard analyses, including using group data aggregated over all pigeons, or individual data that measure the behavior of each pigeon. We then show how an extension to a hierarchical model, which includes both a group and an individual representation of behavior, leads to a more complete account. We demonstrate that the hierarchical model also has the advantage of allowing sensible predictions about behavior in the task in general, beyond the six pigeons observed in the experiment. In the second modeling analysis, we consider the relationship between the standard response ratio and time ratio measures. We find evidence that one part of the matching behavior, related to the sensitivity parameter in the generalized matching law, is the same between these measures for all pigeons. This suggests a theoretical invariance that simplifies the overall model. In the final modeling analysis, we consider a more elaborate model of individual differences. This model relies on a latent mixture structure, which provides a way of allowing for qualitative individual differences, consistent with the presence of sub-groups. We find evidence that there may be at least two sub-groups of pigeons in our data, defined by
fundamentally different bias parameters in the generalized matching law.

Cumulatively, the three analyses in the case study demonstrate many of the major advantages of using Bayesian methods to implement, analyze, and apply behavioral models. In the discussion, we consider the general applicability of these features to models, experiments, and psychological phenomena throughout the experimental analysis of behavior.

The Generalized Matching Law

A major impetus for the use of mathematical modeling in the experimental analysis of behavior was the very influential paper by Herrnstein (1961). He reported that when pigeons have the opportunity to respond to one of two available keys, each delivering reinforcers according to independent variable-interval schedules of reinforcement, the relative rate of responses equals the relative rate of the reinforcers associated with them. This empirical relation became known as the “matching law” and has been widely replicated with different species and responses, and different measures of behavior, including time allocation (Davison & McCarthy, 1988; Williams, 1988).

Almost immediately, however, there were reports of deviation from matching, especially when responses are reinforced with different schedules of reinforcement (LaBounty & Reynolds, 1973). Baum (1974) argued that to understand the deviations from the original matching law it is more useful to work with the data as ratios rather than relative frequencies (see also Staddon, 1968). In particular, Baum (1974) proposed analyzing the relationship between responses and reinforcers in concurrent schedules with the equation

\[ \frac{B_1}{B_2} = \beta_0 \left( \frac{R_1}{R_2} \right)^{\beta_1}, \]  

where \( B_1/B_2 \) represents the ratio of behavior (responses or time allocation) for the two options, and \( R_1/R_2 \) represents the ratio of the reinforcers corresponding to the options. The parameter \( \beta_0 \) has a natural interpretation as a measure of the bias towards the option 1 over option 2, and the parameter \( \beta_1 \) has a natural interpretation as a measure of the sensitivity of the ratio of behavior to changes in the ratio of reinforcers.

Equation 1 became known as the “generalized matching law” and is closely related to the power function proposed by Stevens (1957) in the field of psychophysics. The usefulness of the generalized matching law stems from its ability to distinguish between two different theoretical accounts of deviations from matching. One possibility is that it is caused by a stable bias for one option. Another possibility is that it is caused by a lack of sensitivity to the ratio of reinforcers.

The application of the generalized matching law to data has largely been driven by the
observation that a logarithmic transformation re-expresses the law as
\[
\log \frac{B_1}{B_2} = \beta_1 \log \frac{R_1}{R_2} + \beta_0
\]
which is a simple linear model.\(^1\) As explained in the excellent tutorial provided by Reed and Kaplan (2011), Equation 2 allows for a natural and insightful interpretation of the generalized matching law and its parameters. Geometrically, the sensitivity is the slope of the line relating the log-ratio of behaviors to reinforcers, and bias is the intercept of the line. This formulation allows sensitivity to be interpreted as the change in ratio of reinforcers resulting in a change in the ratio of behaviors, and bias as a fixed tendency in the ratio of behaviors. The special case parameterization \(\beta_0 = 0\) and \(\beta_1 = 1\) represents perfect matching. In terms of deviation from perfect matching, sensitivity values of \(\beta_1 > 1\) represent overmatching, while values of \(\beta_1 < 1\) represent undermatching. In terms of bias, values of \(\beta_0 < 0\) represent a bias toward option 2, while values of \(\beta_0 > 0\) represent a bias toward option 1, which explains the geometric interpretation of sensitivity and bias.

As noted by Killeen (2015), however, the generalized matching law can also be seen as an instance of a logistic function. This perspective opens the possibility of interpreting the model under different possible theoretical frameworks. The first, consistent with the motivations for its development, is its interpretation as a psychophysical model (Stevens, 1957) or as an instance of a random utility model (Killeen, 2015). The second interpretation is as a probabilistic response rule, as an instance of a quantal response model (Goeree, Holt, & Palfrey, 2016). Under this interpretation, the sensitivity and bias parameters of the generalized matching law can be interpreted as balancing between exploration and exploitation (e.g., Daw, O’Doherty, Dayan, Seymour, & Dolan, 2006; Speekenbrink & Konstantinidis, 2015). Indeed, from this perspective, it has an interpretation not far from Herrnstein’s (1970) interpretation of the law of effect as a principle of action.

Whatever the psychological interpretation, the statistical realization that the generalized matching law could be conceived as a linear equation led to the almost exclusive application of classic methods from linear regression (see, for example Baum, 1979; Davison & McCarthy, 1988). This encourages a data analysis perspective on understanding behavior, in which the model is used to make measurements of the properties of data. While Bayesian methods can be productively applied to improve inference based on this “data analysis” approach (Kruschke, 2014), we think the generalized matching law is better understood as a psychological model, formalizing a way in which latent psychological variables control behavioral processes. Bayesian methods are especially well-suited to implementing these sorts of psychological models. Bayesian methods allow theoretically meaningful priors to be used instead of \(\log(\beta_0)\) to allow for a natural geometric interpretation of the parameter.

\(^1\)We use \(\beta_0\) instead of \(\log(\beta_0)\) to allow for a natural geometric interpretation of the parameter.
be placed on the psychologically meaningful parameters. They also allow more flexibility in the assumptions about how data are generated, including assumptions in the form of psychological process models (Lee & Wagenmakers, 2013).

Behavioral Data

Our case study uses new data collected from six pigeons (*Columba livia*), all of which were experimentally naive. The experiment involved concurrent random ratio (RR) and random interval (RI) schedules of reinforcement, following a Findley (1958) procedure in which pigeons could alternate between programs by pecking a center key. In a random ratio schedule, every response the pigeon gives has a fixed probability of being rewarded. For example, in a RR-30 schedule a single response has a probability of $\frac{1}{30}$ of being reinforced. Alternatively, in the RI schedule, there is a fixed probability per second that a reward will be available, and a single response will trigger its delivery. For example, in a RI-30 schedule there is a fixed probability of $\frac{1}{30}$ each second that a reward will be available and delivered after the first response.

In a single experimental session, ten different pairs of RR and RI concurrent schedules were presented. For five of these pairs, the RR schedule was held fixed while varying the values of the RI schedule. For the other five pairs, it was the RI schedule that was held constant. The value of the fixed RR schedule was 30 responses, and the five possible RI values were 7.5, 15, 30, 60, and 120 seconds. For the fixed RI schedule the value was 60 seconds, and the five possible RR schedule values were 15, 30, 45, 60, and 120 responses. An experimental session started by randomly selecting one of the fixed programs, then one pair was randomly sampled without replacement from the five possible combinations of RR RI. The same randomization scheme was then applied to the remaining five pairs. Every experimental session consisted of the presentation of the ten pairs, each lasting for ten reinforcer deliveries followed by a 30 second blackout. Bell and Baum (2017) used a similar procedure in which the fixed schedule was presented as an independent condition. Throughout the case study, we use the data from the last 60 sessions (out of a 110) and the last four reinforcers of each pair, consistent with the empirical evidence that choice behavior is stable for that number of sessions and reinforcers (Baum, 2010).

In this form, our experimental data address a number of fundamental research questions. One regards the optimality or rationality of behavior. Given the experimental design, the optimization of reward predicts a bias toward the RR schedule (Baum, 1981; Rachlin, Battalio, Kagel, & Green, 1981). A second research question raised by our data is whether or not the behavior of the pigeons deviates from perfect matching. This question has been addressed previously in the literature (e.g. Bell & Baum, 2017), but not tested in a rigorous model-based way. A third basic research question is whether the response and
time measures of behavior can be accounted for in the same way. Because the requirements of the schedules are different in terms of response and time behavior, it is natural to ask whether sensitivity and bias differ between the measures. More generally, the major research question is whether the generalized matching law, and the theory it formalizes, provides a good account of behavior.

As always in the empirical sciences, the over-arching goal is to develop, evaluate, and refine models so that they capture the regularities in the data, but also provide the scope to express and understand meaningful variability. Possible sources of meaningful variability include individual differences between the pigeons, the relationship between the psychologically distinct sensitivity and bias parameters, and differences in the behavioral measures. The goal of our case study is to demonstrate the ways in which Bayesian inference can enable the development, evaluation, and applications of candidate models to the behavioral data.

## Bayesian Modeling Analyses

### Group Model

One common approach in the modeling of matching studies is to combine data across individuals, sessions, and reinforcers. This creates a group summary of behavior, which we denote

\[ y_{pm} = \log \left( \frac{B_{1m}}{B_{2m}} \right), \]

where \( B_{1m} \) and \( B_{2m} \) represent behavior in the RR and RI schedules, respectively, for the \( m \)th behavioral measure and the \( p \)th pair of reinforcement schedules in the experiment. There are two possible behavioral measures: the response measure \( m = R \) and the time measure \( m = T \). Following the generalized matching law in Equation 2, we assume these data are distributed as

\[ y_{pm} \sim \text{Gaussian} \left( \beta_0^m + \beta_1^m x_p, \frac{1}{\sigma_m^2} \right) \]

where \( x_p = \log \left( \frac{R_1}{R_2} \right) \) is the logarithm of reinforcement ratio for the \( p \)th pair, aggregated across pigeons, and \( 1/\sigma_m^2 \) is the precision with which the observed data follow the model.\(^2\)

**Graphical model.** In order to apply Bayesian methods, we implement this model as a graphical model. Graphical models were developed and remain widely used in artificial intelligence and machine learning (e.g., Jordan, 2004; Koller, Friedman, Getoor, & Taskar, 2007; Pearl, 1998), and are progressively being adopted in psychology and cognitive science (Farrell & Lewandowsky, 2018; Lee & Wagenmakers, 2013). They are especially well suited

---

\(^2\)Note that we use the standard statistical convention of parameterizing a Gaussian distribution in terms of a mean and a precision. The precision is the reciprocal of the variance, so for a standard deviation \( \sigma \), the variance is \( \sigma^2 \) and the precision is \( \tau = 1/\sigma^2 \). We adopt this convention largely because it is the one used by the JAGS software (Plummer, 2003) we use to implement Bayesian analyses, as provided in the on-line materials accompanying this article.
to the application of computational methods for Bayesian inference. In a graphical model, nodes in a graph represent variables and data, with edges representing the dependencies between them. The basic idea is to provide a probabilistic generative model, formalizing modeling assumptions about how assumed latent parameters control statistical processes to produce the observed data.

Figure 1 shows the graphical model of the generalized matching law applied to group data. The data are represented by $y_{pm}$ node, which is shaded because the data are observed, and circular because the ratios are continuous. The $x_p$ node represents the reinforcement ratio aggregated across pigeons. This node is also shaded, because the ratios are part of the data, and circular because they are continuous. There is an edge from the $x_p$ node to the $y_{pm}$ node in the graph, because an assumption of the model is that the reinforcement ratios influence the observed behavior. The $\beta^1_m$ node represents the sensitivity and the $\beta^0_m$ node represents the bias for the $m$th measure. These nodes are unshaded, because sensitivity and bias are latent model parameters corresponding to psychological variables. The standard deviation is represented by the $\sigma_m$ node. All of these parameters have edges directed to the data, since they all contribute to how the model assumes data are generated.

Graphical model notation uses encompassing rectangles called plates to indicate independent replications of the graph structure within a model. Figure 1 shows two plates. One replicates the sensitivity and bias parameters over the two measures. The other replicates over the pairs of schedules. These two plates intersect for the observed data, which involves all pairs and both measures.

**Specifying priors.** Figure 1 also shows the prior distributions we used for the model parameters. The Bayesian requirement for prior distributions on parameters is often mis-
understood, and is sometimes seen as a drawback or a problematic aspect of the approach. Our view, following Lee and Vanpaemel (2018), is that priors are a necessary and valuable part of specifying a psychological model. In the same way that the likelihood function—that is, the Gaussian distribution based on the generalized matching law—in the graphical model is based on the assumptions about psychological processes, the priors on the sensitivity and bias parameters should be based on assumptions about the psychological variables they represent. These assumptions are probably less easy to formulate, because the relevant theory is less well developed, but the key point is that the basic approach to developing priors is the same creative scientific one that is already used in developing psychological models.

In practice, Lee and Vanpaemel (2018) argue, it is reasonable to develop priors based not only on guiding theory, but on previous values from other relevant data, or from considering the reasonableness of the predictions about data that a choice of priors implies. Accordingly, the priors in Figure 1 allow for sensitivity, bias, and the the standard deviation to take a range of values consistent with theoretical expectations, and previous modeling results. The priors we have chosen do not make strong theoretical commitment, although this is possible, and sometimes desirable, within the Bayesian approach. For example, it would be possible to assume that sensitivity is always positive, so that increases in reinforcers lead to increases in behavior. This assumption could be formalized by constraining the priors on $\beta_m$ to be positive.

**Posterior inference.** We implemented the graphical model in Figure 1, and all of the subsequent graphical models in this article, in JAGS (Plummer, 2003). JAGS is a free and general-purpose scripting language that automates the application of Markov-chain Monte Carlo (MCMC) algorithms for drawing samples from the joint posterior distribution of model parameters. For the graphical model in Figure 1 the sampled parameters are the $\beta_{m0}$, $\beta_{m1}$, and $\sigma_m$ parameters. The joint posterior distribution of these parameters is approximated by the samples, and represents the relative probability that any combination of parameters is the true combination that generated the data, conditional on the assumptions of the model and the observed data.\(^4\)

Formally, the joint posterior distribution is

$$p(\theta | y) \propto p(y | \theta, M) p(\theta, M),$$

\(^3\)Parameterizing the group distribution in terms of a standard deviation, rather than variance or precision, has the modeling advantage of leading to easier semantic interpretation and modeling intuitions for determining theoretically reasonable priors.

\(^4\)For this graphical model, we collected 15,000 samples from each of four independent chains, after a burn-in period of 55,000 discarded samples per chain. We checked for the convergence of the chains using the standard $\hat{R}$ statistic (Brooks & Gelman, 1997), and by visual inspection. The MCMC settings for later graphical models were similar, and were checked in the same way.
where \( \theta = (\beta_0, \beta_1, \sigma^m) \) are the parameters, \( y \) are the observed data, and \( M \) are the modeling assumptions formalized by Figure 1. Equation 3 shows how data are used in Bayesian analysis to update prior probabilities about parameter \( p(\theta) \) to posterior probabilities \( p(\theta | y) \) based on the likelihood of the data under each set of parameter values.

In a traditional analysis of the generalized matching law for these data, single point-estimates for sensitivity and bias would be found using least-squares optimization or maximum likelihood methods. Those points approximately correspond to measures of central tendency—the mean and the mode, respectively—of the full marginal posterior distributions used by Bayesian methods. The marginal posteriors contain additional information about other plausible values for sensitivity and bias. One simple way of summarizing this uncertainty is by the marginal credible intervals for the parameters of the model. A 95% credible interval provides lower and upper bounds such that there is a 95% probability the true value of the parameter lies in the interval. Based on the group data, the highest posterior density 95% credible intervals for the sensitivity parameter \( \beta_1^m \) are (0.10, 0.23) for the response measure and (0.10, 0.31) for the time measure. For the bias parameter \( \beta_0^m \) they are (1.31, 1.49) for the response measure and (0.42, 0.73) for time measure.

**Posterior predictive check.** Parameter inferences, however, are always conditioned on the modeling assumptions, as the explicit presence of \( M \) in Equation 3 makes clear. If those assumptions are unreasonable, then the parameter inferences are not useful. They could be thought of as answering a question that has not been asked. The standard Bayesian method for assessing the descriptive adequacy of a model is through posterior predictive checks (Gelman, Carlin, Stern, & Rubin, 2004). The posterior predictive distribution represents the predictions a model makes about new data \( y' \) based on the posterior distribution of parameters inferred from observing the current data \( y \). Formally, the posterior prediction is

\[
\widehat{p(y' | y)} = \int p(y' | \theta) p(\theta | y) \, d\theta.
\]

Comparing the observed data \( y \) to the posterior predictive distribution it generates provides a basic test of model adequacy. If a model cannot “fit” the available data, it is not descriptively adequate, and further model development is needed.

Figure 2 shows posterior predictions of the group model based on the joint posterior distribution. The posterior predictions are the lines corresponding to the inferred values.

---

5It is worth noting that a confidence interval, from classical statistics, is defined in a very different way from a Bayesian credible interval, and has very different statistical properties (Morey, Hoekstra, Rouder, Lee, & Wagenmakers, 2015).

6There are infinitely many intervals that contain 95% of the posterior density. The highest-probability density credible interval is the unique one that includes parameter values with the greatest density that together make up 95% of the total probability (Kruschke, 2011). This is a common, although not the only, approach to defining credible intervals in Bayesian analysis.
Figure 2. The relationship between the logarithm of reinforcement ratios, on the x-axis, and the logarithm of response and time ratios, on the y-axis. Solid lines represent the posterior predictions of the generalized matching law model based on its inferences from group data. The dotted line represents perfect matching. The markers represent reinforcer pairs. In the left panel the diamond markers show the group behavioral data for the response measure, and circular markers show the group behavioral data for the time measure. In the middle panel, the response measure data are shown for each individual pigeon, distinguished by different colors. In the right panel, the time measure data are shown for each individual pigeon, again distinguished by different colors.

of sensitivity and bias for each behavioral measure. The dotted line represents perfect matching. These two models are shown in all three panels in Figure 2. The panels in Figure 2 differ, however, in the way they treat the behavioral data. The group data on which the inferences are based are shown in the left panel. The diamond markers correspond to the response measure and the circular markers correspond to the time measure. It is clear that the model is able to describe these data well, since the posterior predictive lines capture the observed pattern of change in both response measures with different reinforcement ratios. The middle and right panels of Figure 2, however, show the individual data that the group data summarize. The middle panel corresponds to the response measure and the right panel corresponds to the time measure. For both measures, a different color is used to represent
the behavior of each individual pigeon.

Figure 2 makes it clear that there are significant and interpretable individual differences between pigeons for both measures. For example, there is a pigeon represented by yellow diamonds with a consistently lower response ratio across all pairs of reinforcement schedules. Because of these differences, the posterior predictions provided by the group model do not describe the individual data well. Thus, if the modeling goal is broader than group behavior, and includes accounting for the behavior of individual pigeons, there is a need for further model development.

Individual Model

**Graphical model.** The obvious way to extend the basic model to accommodate individual differences is to allow different pigeons to have different levels of sensitivity and bias. This corresponds to applying the generalized matching law model independently to the individual data for each pigeon. Figure 3 shows a graphical model that has been extended to make this change. It includes a plate that encompasses the individual pigeons, so that there are now independent sensitivity $\beta_{bm}^0$ and bias $\beta_{bm}^1$ for the $b$th pigeon on the $m$th measure. The observed data now take the form $y_{pbm}$ for the $b$th pigeon on the $p$th pair of schedules according to the $m$th measure.

**Posterior inference.** Figure 4 shows the parameter inferences made by the individual model. The left panel relates to the response measure and the right panel relates to the time measure. In each panel, samples from the joint posterior distribution of sensitivity and bias are shown by circular markers, colored according to each individual pigeon. The
The joint posterior distribution of the sensitivity $\beta_1$ and bias $\beta_0$ parameters for the response (left panel) and time (right panel) measures. The main axes in each panel show points from the joint prior distribution in gray, and points from the joint posterior distribution colored according to the individual pigeon. The cross marker shows the point in parameter space corresponding to perfect matching. The distributions on the edge of the main axes correspond to the marginal posterior distribution of the sensitivity and bias parameters, again colored according to the individual pigeon.

Distributions above and to the right show the marginal distributions for bias and sensitivity, again colored according to each individual pigeon. It is clear that the bias towards the random ratio schedule is not the same for all pigeons. In fact, they appear to form three visually distinct clusters. It is less clear, however, that there are individual differences in sensitivity, since the marginal distributions overlap, suggesting that it is possible that all of the pigeons have the same value. Later in the case study, we will pursue these insights, refining the model to capture the invariance in sensitivity, and the sub-group structure of individual differences in bias.

**Posterior predictive check.** In the meantime, Figure 5 shows a posterior predictive check of the individual differences model. The two panels in the left column relate to the individual differences model, with the top panel corresponding to the response measure and the bottom panel corresponding to the time measure. The analysis is the same as used in Figure 2, with the data for each pigeon shown as individual points. There are now six sets of posterior lines, one for each pigeon, relating reinforcement ratios to the behavioral measures, with different colors corresponding to different pigeons. It is clear that allowing for individual differences in sensitivity and bias makes the generalized matching law model
Figure 5. Posterior predictive analysis for the individual model (left column) and group-and-individual model (right column), for the response measure (top row) and time measure (bottom row). Each panel shows the posterior predictions of the generalized matching law relating reinforcement ratios on the x-axis to behavior on the y-axis. The model posterior predictions are shown as lines and the behavioral data are shown as markers, both colored according to the individual pigeons. The dotted line represents the predictions of perfect matching.

Model comparison with perfect matching. Given the descriptive adequacy of the individual model, the inferences about model parameters in Figure 4 are meaningful. For example, they allow us to address the basic question of whether the pigeons show behavior consistent with perfect matching. Statistically, this is a question of model selection rather than parameter inference. The Bayesian approach to model selection is based on the Bayes factor, which is a ratio expressing the evidence data provide for one model over another (Kass & Raftery, 1995). The Bayes factor can be thought of as the change from prior model
odds to posterior model odds, as in the following equation:

$$\frac{\text{posterior odds}}{\text{prior odds}} = \frac{p(M_m | y)}{p(M_g | y)} = \frac{\text{Bayes factor}}{p(y | M_m)} \times \frac{p(M_m)}{p(M_g)}.$$  

(5)

In Equation 5, $M_m$ represents the matching model, which assumes unit sensitivity $\beta_1 = 1$ and no bias $\beta_0 = 0$, while $M_g$ represents the general model with priors on sensitivity and bias specified in the graphical model in Figure 3. The Bayes factor measures how likely each model is to have generated the observed data $y$, averaged over each possible parameterization of the model. Because the Bayes factor is an odds ratio, of the same type as likelihood ratio tests, it has a natural scale for the interpretation of significance that is calibrated by betting. A Bayes factor of 3, for example, means that the data were three times better predicted by one model than the other. Several interpretative scales for Bayes factors have been suggested, associating verbal labels such as “weak evidence” and “strong evidence” to ranges of values (Burnham & Anderson, 2003; Jeffreys, 1961; Kass & Raftery, 1995; Royall, 2000).

It is often convenient to express the Bayes factor in terms of the model that is favored by the data. If the matching model is favored, this Bayes factor is $BF_{mg} = p(y | M_m) / p(y | M_g)$. If the general model is favored, the Bayes factor is $BF_{gm} = p(y | M_g) / p(y | M_m)$. These two Bayes factors are, of course, simply related by $BF_{mg} = 1 / BF_{gm}$, and contain the same information. Our approach is to report the Bayes factor that is greater than one, so that the evidence provided by data are always expressed in terms of the supported model.

One advantage of the Bayes factor is that it is sensitive to all of the components of model complexity (Myung & Pitt, 1997; Pitt, Myung, & Zhang, 2002). Averaging to form the marginal densities achieves this complexity control, because complicated models that predict many possible data patterns are unable to give high predictive weight to the particular data observed (Pitt et al., 2002; Wagenmakers, Morey, & Lee, 2016). This state of affairs contrasts favorably with equating the complexity of a model with a count of the number of parameters in a model. This is the simplifying assumption made by information criteria, such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), which have been applied to model comparisons involving the generalized matching law (Klapes, Riley, & McDowell, 2018; McLean, Grace, & Nevin, 2012). While parameter counts are sometimes a reasonable approximation to model complexity, they are more often inadequate, because they ignore the functional form of parametric interaction. This becomes a serious weakness for the types of hierarchical and latent mixture models we

---

A useful “cheat sheet” summarizing many of these scales is provided by the blog post [https://www.nicebread.de/grades-of-evidence-a-cheat-sheet/](https://www.nicebread.de/grades-of-evidence-a-cheat-sheet/).
consider later in the case study. In latent mixture models, some parameters add much more additional complexity than others. For hierarchical models, it is often the case that increasing the number of parameters decreases the complexity of a model, because the hierarchical structure adds additional constraining theory that is formalized by assumptions about how parameters co-vary within the model (Lee & Vanpaemel, 2018). Even in non-hierarchical settings, it can be the case that adding a free parameter to a model makes the model simpler (see Lee & Vanpaemel, 2018, for a concrete example). The Bayes factor also has the advantage of being independent of the priors placed on the models, simply expressing how those priors should be updated to obtain posteriors.

While Bayes factors have the right statistical properties, they can be difficult to calculate, even with modern computational methods. Fortunately, in this case, it is possible to use the Savage-Dickey density ratio method, which is a scalable computational approach that can be applied to nested model comparisons (Wetzels, Grasman, & Wagenmakers, 2010). An explanation of the Savage-Dickey method in the context of psychological modeling is provided by Lee and Wagenmakers (2013, Chapter 7) and Wagenmakers et al. (2016). The perfect matching model is a special case of the more general model with \( \beta_1 = 1 \) and \( \beta_0 = 0 \). This means that the Bayes factor in favor of the more general model is found by the ratio of the prior to posterior density at this critical point in the parameter space. Figure 4 shows samples from the prior distribution by gray circles. It is clear that the critical point \((\beta_0, \beta_1) = (0, 1)\) has some density under the prior distribution, but negligible density under the posterior distributions for all of the pigeons. Intuitively, the data provide evidence, for every pigeon, that their behavior is not well described by the parameterization that corresponds to perfect matching. This leads to large Bayes factors in favor of the general model. We approximated these Bayes factors numerically, and found them all in excess of 10,000 in favor of the general model (i.e., \( BF_{gm} > 10,000 \)). Thus, the data provide compelling evidence that none of the individual pigeons adheres to perfect matching. This finding speaks directly to the basic theoretical distinction between matching and optimization behavior (Baum, 1981).

**Group-and-Individual Model**

While the individual model proved to be descriptively adequate where the group model was not, the individual model is limited. It is effectively only a model of the specific six pigeons in the experiment. The potential advantage of a group model is that it allows for predictions and generalizations beyond the available data. Hierarchical models provide a

---

8The Bayes factor obviously is sensitive, however, to the prior distributions on model parameters. These priors are what are averaged over to form the marginal densities that define the Bayes factor. For example, for the general model in Equation 5, the marginal density is \( p(y \mid M_g) = \int p(y \mid \beta_0, \beta) p(\beta_0, \beta_1 \mid M_g) \, d(\beta_0, \beta_1) \).
Figure 6. Graphical representation of a hierarchical version of the generalized matching law incorporating individual differences over sensitivity and bias.

mechanism for modeling at both the individual and group levels simultaneously (Shiffrin, Lee, Kim, & Wagenmakers, 2008).

**Graphical model.** The basic idea of a hierarchical model for individual differences is to account for both the group and individual behavior in one model, by making assumptions about how the two relate to each other. Figure 6 shows a graphical model that extends the generalized matching law in this way, to produce a combined group-and-individual model. There are still individual sensitivity and bias parameters for each pigeon but, instead of being independent of one another, they are now assumed to be drawn from a common group distribution. The assumption about the structure of individual differences is that variation in both parameters follows a Gaussian distribution. Thus, the sensitivity of the $b$th pigeon on the $m$th measure is $\beta_1 bm \sim \text{Gaussian}(\mu_{\beta_m}, 1/(\sigma_{\beta_m}^2))$, where $\mu_{\beta_m}$ is the group mean for sensitivity, and $\sigma_{\beta_m}$ is the group standard deviation for sensitivity. The group and individual levels for bias are modeled in the same way.

**Posterior predictive check.** The posterior predictive analysis for the hierarchical group-and-individual model is shown in the right column of Figure 5. The side-by-side comparison makes it clear that the inferences about the sensitivity and bias of individual pigeons are extremely similar to the individual model. This means that the group-and-individual model is descriptively adequate in the same way as the individual model.

**Generalization and prediction.** The main benefit of the group-and-individual model lies in its ability to generalize and make predictions about the behavior of other pigeons. Figure 7 shows the predictions each of the group, individual, and group-and-individual models make about a new pigeon for both behavioral measures. For the group model, the
Figure 7. Predictions of the group model (left column), individual model (middle column) and group-and-individual hierarchical model (right column) for a new pigeon. The top row corresponds to the response measure and the bottom row corresponds to the time measure. Each panel shows the posterior predictions of the model as lines. The individual behavioral data are shown for the hierarchical model.

Predictions about behavior shown in the left column simply follow the inferred sensitivity and bias parameters. Since the group model (implicitly) assumes there are no individual differences, the predictions for a new pigeon are the same as the inferences for the existing pigeons. As noted earlier, the assumption of no individual differences is too strong, and the model fails a basic test of descriptive adequacy.

For the individual model, the prediction about a new pigeon follows the prior assumptions about sensitivity and bias. This is because the model assumes each pigeon is independent of each other, and thus none of the inferences for the six observed pigeons inform predictions about a new pigeon. The resulting predictions, shown in in the middle column of Figure 7 are extremely broad, and essentially allow for any possible linear relationship between reinforcers and behavior ratios. One way of understanding this result is that the individual model is too complicated, allowing for arbitrarily different values of sensitivity and bias for different pigeons, following from the assumption of independence.
The group-and-individual model, however, balances the simplicity of group regularities with the complexity of individual differences. The predictions it makes about a new pigeon correspond to possible values of sensitivity and bias drawn from the group distributions, which, in turn, are inferred from the observed behavior of the six pigeons. As the results in the right column of Figure 7 show, these predictions capture both the regularities across the observed pigeons and the variability between them.

The descriptive adequacy and predictive capabilities of hierarchical modeling are well demonstrated by this example, and highlight the usefulness of the approach in modeling group and individual behavior simultaneously. For this reason, the hierarchical approach is now widely used in many areas of cognitive modeling (e.g., Lee & Newell, 2011; Oravecz, Anders, & Batchelder, 2015; Rouder & Lu, 2005; Rouder, Sun, Speckman, Lu, & Zhou, 2003). The ability to model behavior at multiple levels of analysis seems similarly likely to be useful in many areas in the experimental analysis of behavior. Underscoring this promise, there are several recently-developed hierarchical or multi-level models in the analysis of delay discounting and inter-temporal choice behavior (Chávez, Villalobos, Baroja, & Bouzas, 2017; Vincent, 2016; Young, 2018).

Invariance in Sensitivity

One of the insights from the individual model, apparent in the posterior distributions shown in Figure 4, is that bias seems to change across behavioral measures, but sensitivity may not. That is, there is a suggestion that the sensitivity an individual pigeon has for the response measure is the same as for the time measure. Testing for this equivalence involves comparing a model that assumes the same sensitivity against one that assumes different sensitivities for each measure.

**Graphical model.** Figure 8 presents a graphical model for performing this model comparison, and calculating the required Bayes factors. It extends the individual model in Figure 3 by specifying the relationship between model parameters across the two behavioral measures. Formally, the sensitivity of the bth pigeon on the response measure, $\beta_{bR}$, is defined as $\beta_{bR} = \mu_{b} + \delta_{b} / 2$, while its sensitivity on the time measure, $\beta_{bT}$, is defined as $\beta_{bT} = \mu_{b} - \delta_{b} / 2$. Intuitively, the new $\mu_{b}$ parameter corresponds to the average sensitivity of the bth pigeon across both measures, and $\delta_{b}$ is the effect size of the difference in sensitivity between the two measures. The graphical model in Figure 8 uses exactly the same approach to characterize the difference in the bias parameter across measures for each pigeon.

The rationale for expressing the difference in parameters across measures in this way is to allow the Savage-Dickey method to be used to estimate the required Bayes factors. If, for example, $\delta_{b} = 0$, this means that there is no difference in sensitivity between the measures,
**Figure 8.** Graphical model for testing the equivalence of the sensitivity and bias parameters across behavioral measures, for each individual pigeon.

corresponding to the “null hypothesis” of equivalence. This model is nested within the more general model in which there is a non-zero effect, corresponding to the “alternative model” of a difference in sensitivity across measures for the \( b \)th pigeon. Thus, the ratio of prior and posterior distributions of the effect size parameters \( \delta_{b0} \) and \( \delta_{b1} \) provides Bayes factors for assessing whether sensitivity and bias, respectively, change over behavioral measures for each pigeon.

**Model selection test of invariance.** Figure 9 shows the Savage-Dickey analyses for estimating Bayes factors, for both sensitivity and bias, and for each pigeon. For the bias parameter, there is always overwhelming evidence in favor of there being a difference between the behavioral measures. The posterior for the effect size parameter gives almost no density to the value 0. This is emphasized in the far right panel in the top row of Figure 9, which expands the effect size \( x \)-axis to focus on values near 0. Here, it is clear that the prior, shown by the broken line, gives much greater density to 0 than the six posteriors for the six pigeons shown by the solid lines. Accordingly, five of the six pigeons have Bayes factors \( BF_{ds} \) greater than 1000 in favor of a difference in bias across the measures. One pigeon has a smaller, but still overwhelming, Bayes factor of about 392 in favor of a difference in bias.

For the sensitivity parameter the Savage-Dickey analyses, as shown in the bottom row of Figure 9, reach the opposite conclusion. For every pigeon, there is more posterior density than prior density at an effect size of zero. Thus, the Bayes factors \( BF_{sd} \) favor the null model that the same pigeon has the same sensitivity for both behavioral measures. As Figure 9 shows, the estimated Bayes factors range from 3 to 12. These Bayes factors provide evidence in favor of an equivalence or invariance of sensitivity across the response and time behavioral
Figure 9. Savage-Dickey analyses for estimating Bayes factors that test the equivalence across behavioral measures for the bias and sensitivity parameters. The top row corresponds to bias and the bottom row corresponds to sensitivity. The six columns correspond to the six pigeons. Each panel shows the prior on the effect size as a broken line and the posterior as a solid line. The critical point with effect size zero is indicated by a circular marker. The far right panel in the top column shows an expanded view near the critical point for the bias parameter.

measures.

The analyses based on the graphical model in Figure 9 provide a statistically rigorous approach, based on model comparison, to confirm the intuitions from the parameter inferences in Figure 4. The behavioral data provide evidence for sensitivity being invariant across behavioral measures for a given pigeon, but for bias changing between the response and time measures. This result has a natural theoretical interpretation. Given that reinforcers are correlated with responses in one of the schedules, it is reasonable to expect a higher bias for the response measure than for time. There is, however, no theoretical reason that the sensitivity between measures should be different.

Sub-groups of Bias

A second insight from the parameter inferences in Figure 4 relates to the possible structure of individual differences for the bias parameter. In particular, it appears that there may be up to three qualitatively different sub-groups of pigeons for both of the behavioral measures. While the group-and-individual model in Figure 5 allowed for continuous individual differences, under the assumption that the bias of each pigeon came from a uni-modal Gaussian distribution, it did not allow for this sort of sub-group structure. Qualitative individual differences of this type can be incorporated using latent mixture models, also sometimes known as latent class models (Lee, 2011, 2018). The basic idea is to introduce a
discrete parameter that indicates which of several possible group distributions an individual bias parameter is drawn from. When combined with hierarchical modeling, the latent group approach allows for a very flexible and general account of individual differences, capturing both qualitative differences and quantitative variation (Bartlema, Lee, Wetzels, & Vanpaemel, 2014).

**Graphical model.** Figure 10 shows a graphical model that incorporates a hierarchical latent mixture structure for the bias parameter. A total of $g$ different possible components are included in the mixture, corresponding to the sub-groups of individual differences. These components are defined by different group means $\mu_{0m}^g$ for the $m$th measure. A discrete latent parameter $z_b = 1, 2, \ldots, g$ indicates to which of these groups the $b$th pigeon belongs. Thus, the bias for the $b$th pigeon on the $m$th measure is generated as $\beta_{0bm} \sim \text{Gaussian}(\mu_{0m}^g, 1/(\sigma_{0m})^2)$ where the standard deviation $\sigma_{0m}$ is assumed to be the same for each sub-group. The latent indicator variable is drawn from a base-rate $\phi = (\phi_1, \ldots, \phi_g)$, with $\phi_k$ being that proportion of pigeons belonging to the $k$th group. The indicator variable is drawn according to these proportions $z_b \sim \text{Categorical}(\phi)$, which allows an inference to be made about what proportion of pigeons belong to each sub-group. The base-rate is given a $\phi \sim \text{Dirichlet}(1, \ldots, 1)$ prior, corresponding to the assumption that all possible combinations of proportions in the base-rate are equally likely. To make the model identified, an order restriction is placed on the proportions in the base-rate, so that $\phi_1 \leq \cdots \leq \phi_g$. 

![Graphical representation of the latent mixture model allowing for multiple different sub-groups of bias.](image)
The graphical model in Figure 10 makes assumptions about sensitivity that follow from the previous analyses, which produced the result that sensitivity is invariant across the two behavioral measures. Thus, the model uses only one sensitivity parameter $\beta_1^b$ for the $b$th pigeon, and it is assumed to generate behavior for both response and time measures. The individual differences in sensitivity continue to be modeled hierarchically by a single Gaussian group distribution, as in the original group-and-individual hierarchical model in Figure 6.

**Testing for sub-groups.** We applied the graphical model in Figure 10 to consider $g = 3$ maximum possible sub-groups. Nested within this model are the more limited possibilities that bias is best described as having just one group, or has only two sub-groups. A strict definition would equate the presence of only one group with $\phi_1 = \phi_2 = 0$ and the presence of two sub-groups with $\phi_1 = 0$. We decided to consider a different operational definition, in which a sub-group is only meaningful if it captures at least 1% of the population. Under this definition, the simpler nested hypothesis corresponds to a range in parameter space. The Savage-Dickey method generalizes to areas using what is often called the encompassing priors approach (Klugkist, Kato, & Hoijtink, 2005).

Figure 11 presents the Savage-Dickey analyses for estimating Bayes factors to compare the one group, two sub-group, and three sub-group models, using the 1% operational definition of a meaningful sub-group. The top panels show the marginal distributions for the base-rates of the three possible sub-groups, with the prior shown as a broken line and the posterior shown as a solid line. The bottom panels show the Savage-Dickey analysis for the joint distribution of $\phi_1$ and $\phi_2$. Both of these base-rates being less than 0.01 corresponds to neither the second or third sub-group existing, and so is effectively a single group model. This region of parameter space is highlighted in gray, and the Savage-Dickey analysis estimates a Bayes factor of 2.3 against the single group model. Thus, the data provide some evidence, although it is at most suggestive, that there may be at least two sub-groups. The possibility that there are three sub-groups is tested by the marginal distribution of $\phi_1$ shown in the top row. Here, the Savage-Dickey analysis estimates a Bayes factor of 1.6 in favor of the sub-group not existing.

Thus, overall, it is most likely there are two sub-groups, but it remains possible that there is only one group, or that there are three sub-groups. The posterior distribution of the $z_b$ latent indicator parameters provide the inference, under each of these possibilities, about the sub-group to which each pigeon belongs. The most likely two sub-group model separates the pigeon shown in yellow in Figure 4 from the others. The three sub-group model further separates the pigeons shown in orange and red from those shown in blue and green. We emphasize, however, that Bayes factors of 2.3 and 1.6 provide extremely weak evidence, and the safe scientific conclusion would simply be that the data do not provide enough
Figure 11. Analysis of the posterior distribution of the base-rate $\phi$. The top row shows the marginal posterior distribution (solid line) and prior distribution (broken line) for the proportion of pigeons belonging to each of three possible groups. For $\phi_1$ the Savage-Dickey analysis leading to a Bayes factor 1.6 in favor of $0 \leq \phi_1 \leq 0.01$ is also shown. The bottom row shows the Savage-Dickey analysis of the joint distribution of $\phi_1$ and $\phi_2$, with the prior shown in one panel and the posterior in the other. The critical range in which $0 \leq \phi_1 \leq 0.01$ and $0 \leq \phi_2 \leq 0.01$ is highlighted, and leads to a Bayes factor of 2.3 in favor of the alternative.

Discussion

Our case study provides a number of examples of the two general and complementary ways in which we think Bayesian methods can improve psychological modeling. The first advantage involves basic statistical properties, while the second involves flexibility in model development. We discuss each in turn.

Bayesian statistical advantages

Representing uncertainty. At the heart of updating prior distributions to posterior distributions as new information becomes available is Bayes rule, which follows directly from
the basic axioms of probability theory (Cox, 1961). This principled statistical foundation means that Bayesian methods have a number of desirable properties (Jaynes, 2003; Lindley, 1972). For example, as we noted earlier, Bayesian credible intervals represent uncertainty in a naturally interpretable way that classical confidence intervals do not (Morey et al., 2015). In fact, many researchers misinterpret confidence intervals as credible intervals, suggesting that the Bayesian approach is intuitive (Hoekstra, Morey, Rouder, & Wagenmakers, 2014).

A good example in our case study of the usefulness of representing uncertainty is provided by the joint posterior distribution for the individual model shown in Figure 4. This Bayesian inference shows the distribution of possible combinations of sensitivity and bias, for all six pigeons, that follow from their data and the assumption that the generalized matching law describes their behavior. We think this is exactly the sort of information an empirical scientist wants. It helps for understanding the underlying sensitivity and bias of the pigeons, and for exploring the possibility of invariances of structured differences within and between pigeons. In short, it provides a way of gaining insight into the underlying psychological processes in a way that represents what is known from the model and data, but also acknowledges what remains unknown by representing the uncertainty associated with the inferences.

Evaluating models. A second statistical feature of Bayesian methods is that they make it possible to quantify the evidence data provide for or against any model, including null models that express sameness or invariance, in a way that automatically accounts for both goodness-of-fit and model complexity (Gallistel, 2009; Pitt et al., 2002).

Our case study involved at least two examples of using this capability to address theoretical questions. One involved the question of whether or not the pigeons showed behavior consistent with perfect matching. The Bayes factors that answered this model evaluation question quantified strong evidence against the perfect matching model. When dealing with RR-RI schedules of reinforcement, the standard finding is that matching is not found (Bell & Baum, 2017). In general, however, these findings are not based on model comparison metrics as rigorous as the Bayes factor, with its ability to compare model predictions against data in a way that balances goodness-of-fit and model complexity.

A second use of Bayesian model selection involved testing an open research question in the field, which asks whether sensitivity and bias are invariant with respect to the response and time measures. As summarized in Figure 9, the Bayes factors again answered these theoretical questions decisively. There is compelling evidence that sensitivity is invariant, but bias is not. It is interesting that our conclusions differ from those reached by Baum (1979), in a review of the matching literature, who found sensitivity to be slightly higher for the time allocation measure.

Finally, as was mentioned earlier, the latent mixture model analysis provides only weak
evidence in favor of two sub-groups for the bias parameter. This result is similar to one that can be observed in data provided by Bell and Baum (2017, see Figure A2), in which the least-squares estimate of the bias parameter for each pigeon seems to form at least two distinct groups.

**General applicability.** A third basic statistical feature of Bayesian methods is their ability to function correctly even for small data samples. Bayesian methods have the property that they apply in exactly the same way, regardless of the number of data available (Lee & Wagenmakers, 2005). Our case study considered examples where limited data raised challenges. For example, the attempt to make inferences about structured sub-groups with different bias was based on only six pigeons. Given the limited information available, Bayesian methods behaved in a reasonable way. The Bayes factors summarized in Figure 11 justify only weak and suggestive results, consistent with the lack of available evidence. The ability of Bayes factors to quantify evidence for an alternative model, quantify evidence for a null model, or quantify a result that simply says there is not much evidence for or against either model, is an intuitive and useful set of possibilities.

Bayesian methods have related strengths in dealing with missing data, or data whose measurement is impoverished in some way, such as through censoring. Bayesian statistics is especially good at dealing with missing data, because it automatically infers missing data in a model-based way, based on the available data. Our case study did not consider these features, but Lee and Wagenmakers (2013, Chapter 6) provide examples of applying cognitive models using graphical modeling methods to missing data, and Okada, Vandekerckhove, and Lee (2018) provides examples dealing with censored data.

**Bayesian modeling advantages**

The second advantage of Bayesian methods involves the flexibility that they provide for model development, evaluation, and comparison. The traditional analysis of our data using the generalized matching law would be restricted to the group model or individual model. These are obvious starting points, but are not rich enough to explore all of interesting research questions. Most modeling challenges in psychology require more than a simple account of how a few psychological variables generate one sort of behavioral data. Good theories are more ambitious, and often involve multiple different sorts of psychological variables that interact in different ways, and control complicated psychological processes that affect different sorts of behaviors. Formal models that capture the richness of these sorts of theories are often difficult to analyze using classical statistical methods, but can be handled in the principled way by Bayesian methods. As Lee (2011, 2018) argues, this flexibility frees researchers to develop models creatively, unconstrained by limitations in statistical methods.
Our case study relies heavily on the flexibility and expressive power of Bayesian methods. We started with a simple psychological model, in the form of the generalized matching law, applied to aggregated data. As we tried to develop a better account of the data based on this law, we expanded the model to allow for individual differences, and then to allow for structure individual differences in a hierarchical model.

In a different strand of our model development, we considered the possibility that the same underlying psychological variable of sensitivity simultaneously controlled behavior both in terms of response counts and time allocation. This sort of model structure is sometimes known as a common cause model, and is basic to the empirical sciences. As Lee (2018) points out, one of the reasons to believe in models in physics is that latent variables like the mass of an object influence multiple physical phenomena, like momentum, charge, and gravitational forces. Common cause models provide the capability to test theoretical breadth in psychological phenomena, such as the invariance we found for sensitivity across behavioral measures.

In the final part of our case study, we considered the possibility of discrete or qualitative individual differences using a latent mixture model. This type of model structure makes it possible to consider a range of theoretical ideas in which different psychological processes combine to produce the overall observed data. Latent mixture models have been used, for example, to identify different decision strategies within and between people (Heck, Hilbig, & Moshagen, 2017; Lee, 2016), to account for individual variability in temporal discounting rates (Chávez et al., 2017) to identify contaminant participants in an experiment (Zeigenfuse & Lee, 2010), and to understand how multiple semantic topics interact to produce a text document (Griffiths, Steyvers, & Tenenbaum, 2007). Our use of latent mixture models allowed us to pursue a data-driven insight about the sub-group structure of bias in an exploratory way.

Collectively, as Lee (2011, 2018) argues, hierarchical, common cause, and latent mixture model structures allow for enormous flexibility in developing models of psychological processes. They allow for models to be developed with greater theoretical depth, in which the relationships between psychological variables at different levels of abstraction are modeled hierarchically. They allow for models to be developed with greater empirical breadth, by using common cause structures to model how latent psychological variables and processes influence multiple sorts of behavior in multiple tasks and environments. They allow for richly structured psychological variability to be expressed in hierarchical or latent mixture structures. Simultaneously, they allow for constrained simpler models to be tested against the more ambitious ones, and valued for their simplicity.

The same Bayesian capabilities could be used to expand the account of behavior provided by the generalized matching law itself. To apply the law, we relied upon aggregated data,
consistent with a standard “molar” level of analysis in the field (Shimp, 2014). The alternative is a more detailed and process-oriented “molecular” level of analysis, in which individual behaviors are modeled, rather than aggregated over time. Modeling all 110 sessions, rather than just the last 60, would require incorporating a theory of learning. Modeling all of the reinforcers, rather than just the last four in each pair, would require incorporating a theory of adaptation. It might be that additional psychological phenomena, like sequential effects, would also need to be incorporated to provide a more detailed model at the molecular level of analysis. The Bayesian approach naturally allows these sorts of extended models to be developed and evaluated, using exactly the same methods relied upon in our case study.

In the same way that we used Bayesian methods to provide a more complete account of individual differences and common causes applied to the basic generalized matching law, the methods could be used to make the generalized matching law a more complete account of moment-to-moment behavior.

Conclusion

We hope that our case study has demonstrated in a concrete way how Bayesian methods can be applied to both focused model evaluation and exploratory development, in the quest to understand human and animal behavior. The key methodological point is that the Bayesian statistical principles of inference—invoking prior and posterior distributions over parameters, and prior and posterior predictions about data—provide a coherent, consistent, and interpretable framework for applying psychological models to data, and for developing, evaluating, and applying the models.

As our case studies show, the graphical model formalism and computational methods for inference make this a feasible approach. Thus, it is possible to use Bayesian methods to apply statistical rigor to the evaluation and comparison of psychological models, by always grounding their evaluation in terms of the evidence that is available. The guiding principles of Bayesian methods, based on representing uncertainty and being sensitive to complexity, serve to provide a theoretical freedom not available using traditional modeling methods.

For this reason, the adoption of the Bayesian framework means that theories and models become primary, and statistical methods simply serve in their evaluation and development. We think this is the right set of priorities. There is nothing more limiting than statistical methods that constrain the models and theories that can be considered. We hope and expect that as the experimental analysis of behavior progresses, Bayesian methods will play a key role in permitting better evaluation and application of a wider set of theories and models.
Acknowledgments

Supplementary material associated with this article, including code and data, is available as a project page on the Open Science Framework at osf.io/nyxec/. This research was supported by the projects PAPIIT IG120818 and PRODEP grant UDG-PTC-1369. A preliminary version of this work was presented at the Society for the Quantitative Analysis of Behavior meeting in 2016.

References


