



A simple and flexible Bayesian method for inferring step changes in cognition

Michael D. Lee¹

Published online: 31 July 2018
© Psychonomic Society, Inc. 2018

Abstract

Human behavioral data often show patterns of sudden change over time. Sometimes the causes of these step changes are internal, such as learning curves changing abruptly when a learner implements a new rule. Sometimes the cause is external, such as people's opinions about a topic changing in response to a new relevant event. Detecting change points in sequences of binary data is a basic statistical problem with many existing solutions, but these solutions rarely seem to be used in psychological modeling. We develop a simple and flexible Bayesian approach to modeling step changes in cognition, implemented as a graphical model in JAGS. The model is able to infer how many change points are justified by the data, as well as the locations of the change points. The basic model is also easily extended to include latent-mixture and hierarchical structures, allowing it to be tailored to specific cognitive modeling problems. We demonstrate the adequacy of this basic model by applying it to the classic Lindisfarne scribes problem, and the flexibility of the modeling approach is demonstrated through two new applications. The first involves a latent-mixture model to determine whether individuals learn categories incrementally or in discrete stages. The second involves a hierarchical model of crowd-sourced predictions about the winner of the US National Football League's Most Valuable Player for the 2016–2017 season.

Keywords Step change · Change points · Change detection · Bayesian inference · Spike-and-slab priors

All sorts of human behavior come in the form of binary responses: People choose to press a “thumbs up” or “thumbs down” button, to book a window or an aisle seat, or to stay home for dinner or dine out. The rates of these binary responses usually change over time, and sometimes change suddenly and significantly. A scandal can see the rate of “thumbs up” responses for a celebrity or politician fall dramatically, a bad experience in a window seat can shift the preference to an aisle, and the probability of dining out may jump as the week-end approaches.

As these examples make clear, the statistical problem of detecting step changes in binary responses has applications throughout the cognitive sciences. There is a substantial

statistical literature on methods for detecting change points. Many early methods relied on classic significance testing to find changes, using standard parametric and nonparametric tests (e.g., Pettitt, 1979; Venter & Steel, 1996). Bayesian approaches were also developed relatively early. A. F. M. Smith (1975), for example, provided an analytic Bayesian solution to a number of cases involving binomial and Gaussian distributions with a single change point. More recent Bayesian methods have relied on computational approaches to inference. One approach is to use general methods—such as reversible-jump Markov-chain Monte Carlo (MCMC) methods (Green, 1995), product partition methods (Barry & Hartigan, 1993), discrete-state Markov processes (Chib, 1998), and Gibbs sampling (Stephens, 1994)—and apply them to change-point inference as a special case of the more general capabilities the methods provide. Other computational Bayesian approaches have been specifically developed for change-point inference. For example, Fearnhead (2006) developed a clever recursive approach to MCMC computation that allows efficient inference for large problems, and Adams and MacKay (2007) developed a method for the online detection of change points.

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13428-018-1087-7>) contains supplementary material, which is available to authorized users.

✉ Michael D. Lee
mdlee@uci.edu

¹ Department of Cognitive Sciences, University of California, Irvine, CA, USA

These methods do not seem, however, to be widely used in cognitive modeling. In particular, computational Bayesian methods for the simultaneous inference of the number of change points and their location are rarely used. One barrier to their adoption may be difficulty of implementation, since the methods involve nontrivial algorithms. Another barrier may be the difficulty of adapting or extending published methods to specific psychological modeling problems. Unless the substantive psychological problem takes exactly the same statistical form as the existing solution—which often involves technically convenient choices of simple sampling distributions and conjugate priors—additional theoretical and practical work is required. In this article, we provide a Bayesian statistical approach to the multiple-change-point problem that is designed to help address these barriers. Our approach has the practical attraction of a simple implementation as a graphical model in standard JAGS software (Plummer, 2003). Because of this implementation, the approach also has the theoretical attraction of making it relatively easy to extend and adapt the basic change-point detection model to specific cognitive modeling problems.

The structure of this article is as follows. First, we present the basic statistical approach to inferring change points and its implementation as a graphical model. We highlight the way in which the model combines Bayesian model selection and parameter inference through careful specification of prior distributions. After a simple demonstration with simulated data, we apply the model to the classic Lindisfarne scribes problem (Silvey, 1958), showing that its inferences match those of existing methods. We then demonstrate the flexibility of the modeling approach in two case studies. The first case study involves human category learning data and uses latent-mixture modeling to examine whether learning curves show gradual or step-wise change. The second case study involves crowd-sourced data in which people made predictions about which US National Football League (NFL) players would win the Most Valuable Player (MVP) award. A hierarchical extension to the model is shown to be able to track the current opinion of the crowd and to make an accurate prediction about the winner.

A model for inferring change

Here the basic change-point model is presented, using some simple illustrative data. After describing the data, we develop the statistical model and detail its software implementation. The remainder of the section demonstrates the model by applying it to illustrative data.

Illustrative data

Figure 1 show some illustrative data, involving binary responses for a set of $m = 3$ items made at $n = 20$ discrete

sequential times. For the i th item at the j th time u_{ij} and v_{ij} are the number of responses for the two possibilities, with $t_{ij} = u_{ij} + v_{ij}$ being the total number of responses. Each panel corresponds to an item, with circles showing the observed proportion for one alternative over for that item over times.

The area of the circles is proportional to the total number of responses at that time. These simulated data were designed to capture some basic possibilities in step changes over time. The proportion for item A is always around 0.7; item B changes from 0.3 to 0.7 after the 11th time period; item C starts at 0.9, changes to 0.1 after the 5th time period; then changes to 0.3 after the 14th time period. For all three items, the total number of responses varies across the time periods.

Modeling assumptions

The core modeling assumptions are most easily introduced in the context of a single item with u_j responses for one alternative out of t_j total responses for the j th time. It is assumed that there are $\delta < n$ change points $\tau_1, \dots, \tau_\delta$ that collectively partition the times into $\delta + 1$ stages. The first stage includes all the time periods, $1 \leq j < \tau_1$, the second stage includes all the time periods, $\tau_1 \leq j < \tau_2$, and so on, until the final stage $\tau_\delta \leq j \leq n$. The stage of at the j th time is $w_j = 1, 2, \dots, \delta$ and can be determined by counting the number of change points τ_k exceeded by the j th time, so that $w_j = \sum_k \mathcal{I}(j \geq \tau_k)$, where $\mathcal{I}(\cdot)$ is the indicator function. Once the stage for the at the j th time is determined, the observed responses are simply modeled as $u_j \sim \text{Binomial}(\theta_{w_j}, t_j)$.

The key to our statistical approach to the change-point inference problem is the prior employed for the change point parameters. The number of change points δ and their values $\tau_1, \dots, \tau_\delta$ are able to be inferred jointly by using a spike-and-slab prior (Mitchell & Beauchamp, 1988; Rouder, Haaf, & Vandekerckhove, 2018). These priors are well developed in statistics and are often useful for simultaneous model selection and parameter inference. Specifically, for each potential change point $k = 1, \dots, \gamma$, the model introduces an unordered set of change points τ'_k and assigns the prior

$$\tau'_k \sim \text{Categorical} \left(\frac{n}{2n-1}, \frac{1}{2n-1}, \dots, \frac{1}{2n-1} \right)^{n-1}.$$

This categorical prior places a spike at the value $\tau'_k = 1$ and the slab across the remaining times $\tau'_k = 2, \dots, n$, so that the prior mass given to a change point at the first time is the same as the prior mass given to all subsequent times combined. The insight is that, since inferring a change point at the first time period does not contribute to defining a stage, it corresponds to the inference that the change point does not exist. Thus, there is a prior of one-half for each potential change point that

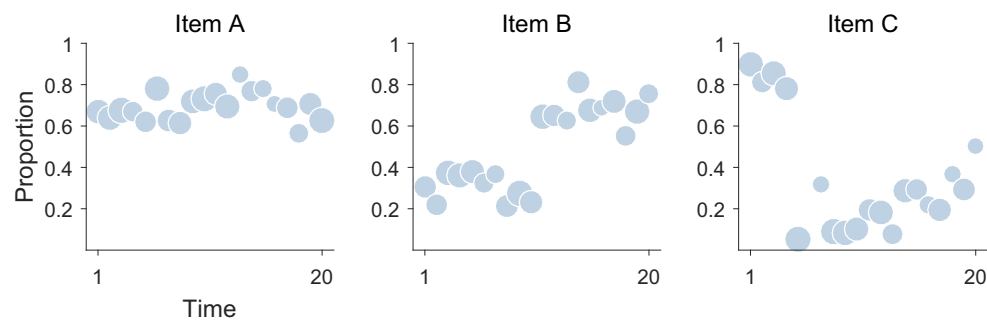


Fig. 1 Illustrative binary opinion data for three items, each measured at 20 times. The circles show the proportions of responses for one alternative over the other for each item at each time, and the area of each circle is proportional to the total number of responses at each time

it does not exist. The equal prior mass given to all subsequent change point then corresponds to assuming that, if a change point exists, it is equally likely to occur at any time. Thus, the prior on τ'_k simultaneously supports the model selection inference about whether or not a change point exists and the parameter inference about the time of the change conditional on it existing.

The value γ is a constant that sets the maximum number of potential change points included in the inference. Conceptually, setting γ formalizes prior assumptions about the number of change points. Specifically, since each τ'_k has prior probability $\frac{1}{2}$ of corresponding to a change point, the prior probability of δ change points follows the binomial distribution $\binom{\gamma}{\delta} \frac{1}{2^\gamma}$. Thus, the maximum prior probability will be given to the number of change points being half of γ , and a good guiding heuristic is to set γ to be about twice as large as the number of change points expected.

The final part of the inference is to define the sorted sequence $\tau_1 \leq \tau_2 \leq \dots \leq \tau_\gamma$ of the τ'_k parameters. This sorting imposes an order constraint to make the model identifiable, so that τ_γ is the last potential change point in the time sequence, $\tau_{\gamma-1}$ is the second-to-last change point, and so on.

Graphical model

A graphical model that follows these modeling assumptions, and applies them to a set of items, is presented in Fig. 2. Graphical models provide an intuitive formalism for representing probabilistic generative models and are well suited to the application of computational methods for Bayesian inference (Jordan, 2004; Koller, Friedman, Getoor, & Taskar, 2007; Lee & Wagenmakers, 2013). In graphical models, the parameters, data, and other variables are represented by nodes in a graph, and dependencies between them are represented by the structure of the graph, with children depending on their parents. Different types of nodes are used to indicate what they represent, and we follow the conventions according

to which observed data or variables are shaded while unobserved parameters are unshaded, continuous values are circular while discrete values are square, and stochastic variables are single-bordered while deterministic variables are double-bordered. In addition, we use encompassing plates to indicate repetition in the graph structure. The graphical model in Fig. 2 has an outer plate for items and an inner plate for times.

The basic version of the model shown in Fig. 2 assumes that each stage has the same rate for one response alternative over the other, but that these rates differ independently across stages. Thus, for the i th item there are change point parameters $\tau_i = (\tau_{i1}, \dots, \tau_{i\gamma})$ and rate parameters $\theta_i = (\theta_{i1}, \dots, \theta_{i,\gamma+1})$ with uniform prior distributions over the range 0 to 1. The relationship between the time j and the change points determine the stage w_{ij} . The response count u_{ij} then depends on the rates, the stage, and the total number of responses t_{ij} . The natural generative interpretation of the model is that the graphical model shows how the observed counts for a given item at a given time are produced by an underlying stage-dependent rate.

Implementation

We implement the graphical model in Fig. 2 using JAGS (Plummer, 2003), which is standard and free software. JAGS automatically applies computational methods for sampling from the joint posterior distribution of a model, allowing for fully Bayesian analysis. Our implementation involves two JAGS scripts, run sequentially, with the output of the first providing input to the second.

The first JAGS script implements the graphical model in Fig. 2 and is listed below. The data are provided in matrices u and t , giving the number of responses for one alternative and the total number of responses for each item and each time period. The other inputs are the number of items m , the number of times n , and the maximum number of possible change points γ .

```
# Basic change detection on votes

model{
  for (i in 1:m){
    for (j in 1:n){
      # response count data
      u[i, j] ~ dbin(theta[i, w[i, j]], t[i, j])
      # stage for time period
      w[i, j] = sum(wTmp[i, j, 1:gamma]) + 1
      for (k in 1:gamma){
        wTmp[i, j, k] = step(j - tau[i, k])
      }
    }
    # prior for rates
    for (k in 1:(gamma+1)){
      theta[i, k] ~ dunif(0, 1)
    }
    # prior for change points
    for (k in 1:gamma){
      tauPrime[i, k] ~ dcat(alpha)
    }
    # order constraint
    tau[i, 1:gamma] = sort(tauPrime[i, 1:gamma])
  }
  # spike-and-slab prior
  alpha[1] = n/(2*n - 1)
  for (k in 2:n){
    alpha[k] = 1/(2*n - 1)
  }
}
```

This script follows the graphical model closely. It uses temporary variables $wTmp$ to calculate the stage for each item at each time period, $tauPrime$ to implement the order constraint on the change points, and $alpha$ to define the spike-and-slab prior for the categorical distribution on the change points.

The joint posterior $\tau_i | \mathbf{u}_i, \mathbf{t}_i$ provides the inference about the number and location of the change points for the i th item. Conceptually, the joint posterior is a probability mass distribution over each possible set of change points. There are many ways this joint posterior could be analyzed, but here we focus on one approach that emphasizes finding an interpretable set of change points.¹ This is done by collapsing the uncertainty in the joint posterior and using the single most likely set of change points. This corresponds to the mode τ_i^* of the joint posterior distribution. Inferences about the rates are then be conditioned on this mode, so that the inferred rates correspond to $\theta_i^* | \tau_i^*$.

¹ If we had different goals, other methods for handling the joint posterior might be appropriate. For example, if our goals focused on the prediction of unseen data, it would be better to use a Bayesian model-averaging approach and to incorporate all the possible sets of change points given nonnegligible mass in the posterior.

It is straightforward to identify the mode of the joint posterior for τ and then to find the posterior distribution for θ from those samples for which τ corresponds to this mode. It is possible, however, that the modal τ would be sampled relatively rarely, which means that the posterior distribution would not be well approximated. Thus, our implementation uses the first JAGS script only to find the posterior mode for τ . That mode is then supplied as an *observed* variable to the second JAGS script, listed below.

```
model{
  for (i in 1:m){
    # Data
    for (j in 1:n){
      # response count data
      u[i, j] ~ dbin(theta[i, w[i, j]], t[i, j])
      # stage for time period
      w[i, j] = sum(wTmp[i, j, 1:gamma]) + 1
      for (k in 1:gamma){
        wTmp[i, j, k] = step(j - tau[i, k])
      }
    }
    # prior for rates
    for (k in 1:(gamma+1)){
      theta[i, k] ~ dunif(0, 1)
    }
  }
}
```

This script is simply a reduced version of the graphical model that assumes τ is known. Accordingly, the inferences for θ are conditioned on the mode for τ .

Application to illustrative data

Figure 3 shows the results of applying the model to the illustrative data. The bottom panels relate to the results of the first step in the implementation and show the joint posterior $\tau_i | \mathbf{u}_i, \mathbf{t}_i$, based on an upper bound of $\gamma = 3$ change points. The resulting three-dimensional joint posterior is shown for each item, with the volume of each cube corresponding to the inferred posterior mass. For item A, the mode is at $\tau_A^* = (1, 1, 1)$, corresponding to the inference that there are no change points. For item B, the mode is at $\tau_B^* = (1, 1, 11)$, corresponding to the inference that there is a single change point at the 11th time period. For item C, the mode is at $\tau_C^* = (1, 5, 14)$, corresponding to the inference that there are change points at the 5th and 14th time periods.

The top panels relate to the results of the second step, in which the rate for each stage is inferred. For each item, the data are again shown by circles, with the inferences about the stages and their

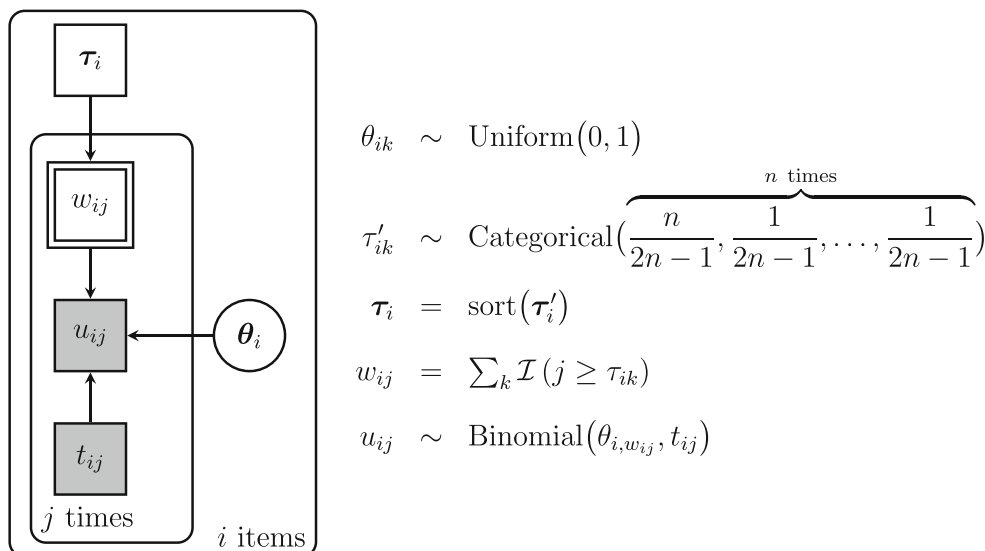


Fig. 2 Graphical model representation of the basic change-point model

rates superimposed by a broken line. These inferred rates correspond to the posterior expectation of the appropriate rate parameter for each stage. For item A there is only one stage, and the rate is always about 0.7. For item B the rate increases from about 0.3 to 0.7 at the change point at the 11th time period. For item C the rate falls from about 0.85 to 0.1 at the 5th time period, then increases to about 0.3 at the 14th time period. These patterns of change are visually reasonable, given the data, and correspond to the way in which the illustrative data were generated.

Application to Lindisfarne scribes’ problem

Blakeley (1949) discusses a philological problem involving the historical use of different *s*-forms for English verbs, such as *drives* and *driveth*. The problem concerns a set of texts known as the Lindisfarne Gospels, which vary in the frequency of their use of these forms. The changes are believed to reflect sequential changes in the scribes of the texts, with each scribe producing some number of consecutive texts before being replaced by the next scribe. Ross (1950) noted that this philological problem can be treated as one of statistical inference, in which the goal is to infer the how many scribes are involved and which texts they are responsible for, as well as the underlying rates of use of the different *s*-forms for each scribe. Silvey (1958) provided an early attempt to apply statistical methods to make these inferences, and A. F. M. Smith (1975) provided an impressive early Bayesian approach.

We consider the same data used by A. F. M. Smith (1975, Table 1), which give counts for each form for a sequence of 13 texts. The results of applying the graphical model in Fig. 2 are shown in Fig. 4. These results are based on

setting $\gamma = 10$ as an upper bound on the number of change points.² Two change points are detected, after the fourth and fifth texts. The overall inference is that a first scribe produced Texts 1–4, a second scribe produced Text 5, and a third scribe produced Texts 6–13. These results, both in terms of the number of change points and their locations, agree with those found by A. F. M. Smith (1975, Tables 4 and 5) using analytic methods.

Although it is comforting that our approach produces results that match analytic results for a simple problem, the real advantage of the graphical model implementation is its flexibility. It is straightforward to extend the basic model for inferring change points beyond the case in which there is one fixed binomial rate for each stage. The following two applications aim to demonstrate these sorts of expanded modeling possibilities, which allow the basic approach to be applied to a broad range of cognitive modeling problems.

Application to category learning curves

In the first application, we consider human category learning performance. The research question is whether learning curves show gradual improvement or step change increases. This inference problem has basic theoretical implications for the roles of two competing accounts of human learning (Ashby & Maddox, 2005; Kruschke, 2008). Gradual improvement is consistent with various theories of associative learning or

² Technical details: The JAGS results for this analysis, and for all of those that follow, are based on six chains, each with 1,000 samples, collected after 5,000 burn-in samples and with thinning every 50th sample. Trace-plots and the standard \hat{R} measure (Brooks & Gelman, 1997) provided good evidence of convergence.

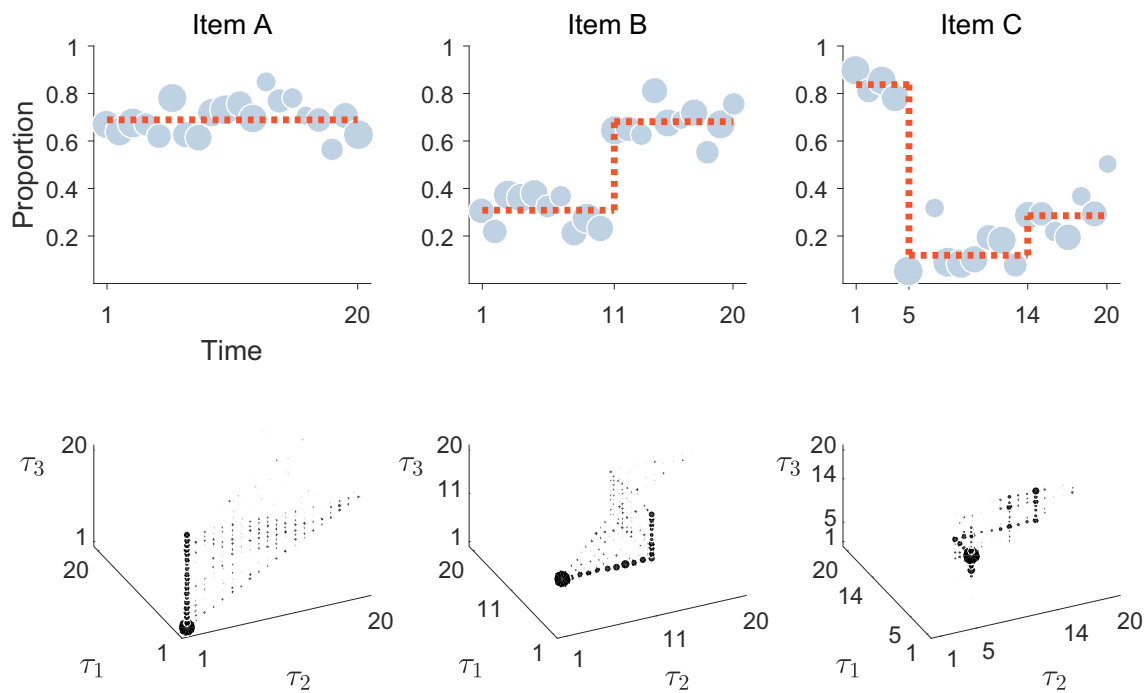


Fig. 3 Results of applying the model to the illustrative data. The top panels show the model inferences for each item. The circles show the response proportions for each item at each time period, and the area of each circle is proportional to the total number of responses for the time

period. The broken line shows the inferred rate over time, including one change point for item B and two change points for item C. The bottom panels show the joint posterior distribution for the τ parameters for each item, the mode of which determines the change points

reinforcement learning, in which small changes are regularly made to update how stimuli are related to categories. Step change increase is consistent with various theories of hypothesis-testing or rule-based learning, in which new schema are used to determine responding, leading to sudden and potentially large improvements in performance.

Data

We use the data from J. D. Smith and Minda (1998, Exp. 3, non-linearly-separable condition) and involved 16 participants learning to categorize eight nonsense words into two categories

based on trial-by-trial feedback. We analyze the learning performance in terms of ten blocks with 56 balanced trials in total, comprising seven presentations of each of the eight stimuli. Thus, the basic data are, for each participant, a sequence of counts over ten blocks of how many correct responses were made out of 56 trials.

Graphical model

Figure 5 shows a latent-mixture model that extends the basic change-point graphical model in Fig. 2. There are two components to the latent mixture. One involves step changes in

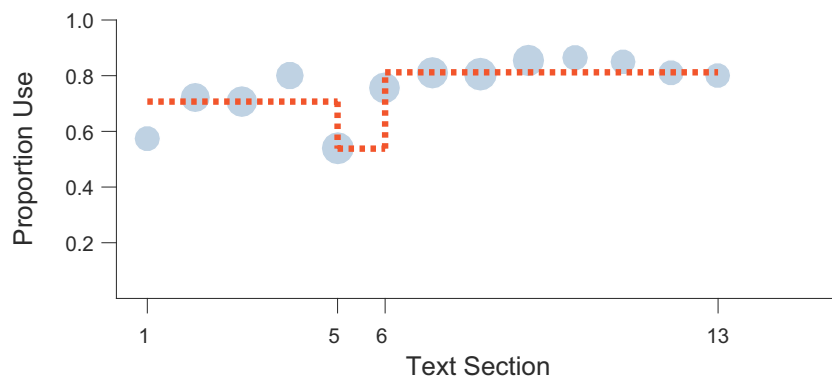


Fig. 4 Results of applying the model to a version of the Lindisfarne scribes data. The circles show the proportions of one variant form over another for each of the 13 sections of the text, and the area of each circle is proportional to the total number of occurrences in that section. The

broken line shows the inferred rate of use of the variant form, including change points, interpreted as changes to different scribes, between Texts 4 and 5, and between Texts 6 and 7

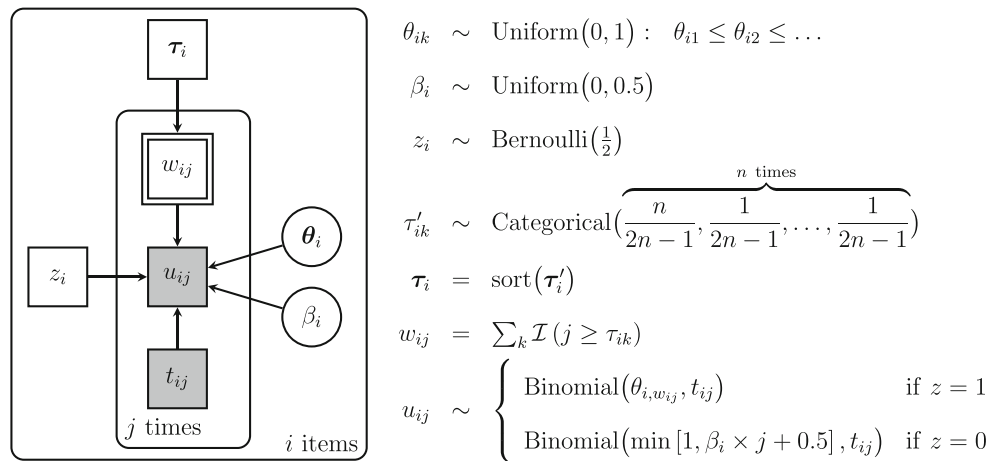


Fig. 5 Graphical model representation of the latent-mixture extension of the change-point model

learning performance, with a sequence of increasing probabilities of correct categorization. The other involves a gradual linear increase in performance, starting at chance-level performance and bounded by perfect performance. The latent binary indicator parameter z_i determines whether the learning curve for

the i th participant is modeled by the step-change or linear-increase mixture component. The step-change component follows the basic change-point model, inferring how many changes there are in the rate of correct responses, at which blocks these changes occur, and what the rate of correct responding is

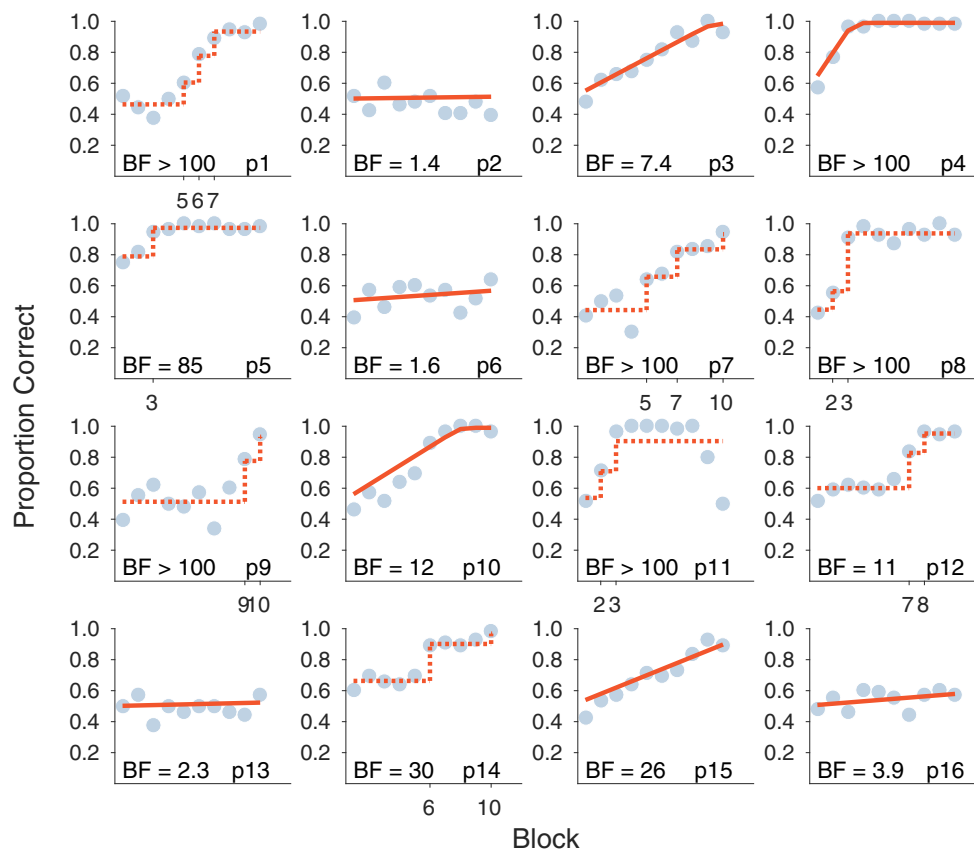


Fig. 6 Results of applying the latent-mixture model to the category learning performance reported by J. D. Smith and Minda (1998, Exp. 3, non-linearly-separable condition). Each panel corresponds to a participant, labeled p1–p16, with circles showing their categorization accuracy over ten blocks of trials. For participants classified as having step-change performance, a broken line shows the inferred rate of

accuracy over each stage, with the x -axis tick labels detailing the change points. For participants classified as having linear-increase performance, a solid line shows the pattern of change implied by the inferred learning rate. Each panel also presents the Bayes factor (BF) in favor of the inferred classification

in each stage defined by the change points. The rates are constrained to increase as the stages progress. The linear-increase component assumes that the rate of correct responding begins at 0.5 and then increases by β_i over each block, until (possibly) the rate of correct responding becomes perfect at 1. The learning rate β_i is specific to the i th participant and is given the prior $\beta_i \sim \text{Uniform}(0, 0.5)$, corresponding to the assumption that performance increases over blocks and can be perfect as early as the second block. The binary indicator variable z_i is given the prior $z_i \sim \text{Bernoulli}(\frac{1}{2})$, corresponding to the assumption that it is equally likely that each participant will belong to the step-change or the linear-mixture component.

The JAGS script below implements the graphical model in Fig. 5. The key changes are that $u[i,j]$ now has a binomial rate that depends on the $z[i]$ latent indicator variable and that a prior is set for each of these indicators in the new #model indicator section of the script. The linear possibility for the rate is implemented in the new #linear component section of the script.

```

model{
  for (i in 1:m){
    for (j in 1:n){
      # response count data follows latent mixture component
      u[i, j] ~ dbin(equals(z[i], 0)*thetaLine[i, j]
                    + equals(z[i], 1)*thetaRate[i, w[i, j]], t[i, j])
      # stage for time period
      w[i, j] = sum(wTmp[i, j, 1:gamma]) + 1
      for (k in 1:gamma){
        wTmp[i, j, k] = step(j - tau[i, k])
      }
    }
    # prior for change points
    for (k in 1:gamma){
      tauPrime[i, k] ~ dcat(alpha)
    }
    # order constraint
    tau[i, 1:gamma] = sort(tauPrime[i, 1:gamma])
  }
  # spike-and-slab prior
  alpha[1] = n/(2*n - 1)
  for (k in 2:n){
    alpha[k] = 1/(2*n - 1)
  }
  # prior for rates
  for (i in 1:m){
    # mean rate for stages
    for (k in 1:(gamma+1)){
      thetaRateTmp[i, k] ~ dunif(0, 1)
    }
    # order constraint
    thetaRate[i, 1:(gamma+1)] = sort(thetaRateTmp[i, 1:(gamma+1)])
  }
  # linear component
  for (i in 1:m){
    beta[i] ~ dunif(0, 0.5)
    for (j in 1:n){
      thetaLine[i, j] = min(0.99, beta[i]*(j-1) + 0.5)
    }
  }
  # prior for model indicator
  for (i in 1:m){
    z[i] ~ dbern(0.5)
  }
}

```

First, we use the full joint posterior inferred by this script to find the modal latent indicator parameters z^* . If this mode is $z_i^* = 1$, the participant is classified in the step-change mixture component, whereas if the mode is $z_i^* = 0$, the participant is classified in the linear-increase mixture component. For participants in the step-change component, the mode of the joint posterior $\tau_i | z_i = 1, \mathbf{u}_i, \mathbf{t}_i$ of their change-point parameters, conditional on assignment to the step-change component, is found. Both the modal z_i^* and τ_i^* for each participant are then provided as observed values for a second script that infers just the θ_i or β_i parameters for each participant, depending on their mixture component classification. The second script is not shown, but simply removes the sections that define priors for the z and τ parameters. It is available as [supplementary material](#).

Modeling results

Figure 6 shows the results of applying the latent-mixture model to the J. D. Smith and Minda (1998) data, with the setting $\gamma = 5$. The observed proportion of correct categorization decisions for each participant over blocks is shown, together with a summary of the inferences of the model for that participant. If the participant is classified as having step-change learning, the model inferences take the form of a series of mean accuracies for the stages, defined by the inferred change points. If the participant is classified as having linear-increase learning, the model inferences are a bounded line based on the learning rate parameter. In each panel, the Bayes factor, derived from the posterior expectation of the z_i indicator parameters, is also shown (Lodewyckx et al., 2011). Thus, for example, the Bayes factor for Participant 14 is 30 in favor of step-change learning, where the participant's accuracy improves from about 65% to about 90% at the sixth block, whereas the Bayes factor for Participant 15 is 26 in favor of linear-increase learning, where this participant's accuracy improves from 50% to about 90% over the blocks.

Overall, Fig. 6 shows exactly half the participants are inferred to use each of the two learning possibilities. Sometimes the certainty of these inferences, as quantified by the Bayes factor, is relatively weak, especially for participants—such as Participants 2, 6, 13, and 16—who show little evidence of learning. This pattern of change in accuracy can be described well by both learning possibilities, but the linear-increase one is generally slightly preferred because it is simpler. For many other participants, however, there is clear and interpretable evidence for either step-change or linear-increase learning. The learning trajectory for Participant 8, for example, shows a clear step-change pattern, with a small increase on the second block and then a large increase on the third block. Similarly, Participant 7 shows a clear step-change pattern. They perform near chance for the first four blocks, improve to about 60% accuracy for the next two blocks, improve again to about 80% accuracy for the next three blocks, and finally reach

near-perfect accuracy after one more increase at the final block. The learning trajectory for Participant 10, on the other hand, shows a steady increase until near-perfect accuracy is reached for the last two blocks. Similarly, Participant 4 is inferred to learn with a linear increase, but with a much greater learning rate, reaching near-perfect performance on the third block.

We also applied the model to the other five conditions presented by J. D. Smith and Minda (1998). The results are available as [supplementary information](#), and continue to show evidence for some participants being better modeled by the step-change account of learning than by the linearly increasing one. We do not attempt to draw conclusions about the general prevalence or theoretical importance of step-changes based on these results. Reaching these conclusions would require considering a wider range of alternative models than the simple linearly increasing account used here. What we have achieved, however, is to demonstrate that an account based on step changes can be incorporated into the analysis of category learning curves using the basic change-point model.

Application to crowd-sourced voting data

In the final application, we consider voting data from the crowd-sourced opinion website www.ranker.com. This site consists of tens of thousands of lists, each containing some number of items. Users can rank the items and can up-vote or down-vote each item at any time. The patterns of voting thus provide a measure of crowd opinion about items as opinion evolves over time. We consider voting data for a sporting award—the Most Valuable Player (MVP) of the US National Football League (NFL)—with the goal of detecting changes in the crowd opinion for different players and using the current opinion for each player at the time the award is determined as a prediction of the winner.

Data

The data come from the ranker.com list “NFL players most likely to be the 2016–2017 MVP.” This list was created on November 18, 2016, and received 31,907 votes for 27 different players, up until Matt Ryan was announced as the winner on February 4, 2017. We consider eight players out of the 27, constituting the leading candidates for the MVP award. These players all received a large number of votes on the Ranker site’s list and include all of the award favorites discussed in the media.

Graphical model

Because the crowd-sourced data involve different people voting at different times, the basic model must be extended hierarchically to allow for variability around a mean group-level rate of

positive opinion. Figure 7 shows the hierarchical graphical model. The mean rate for the x th stage is μ_x , and the standard deviation for the variation around this mean is σ . This means that the rate for the i th item at the j th time is $\theta_{ij} \sim \text{Gaussian}_{(0,1)}(\mu_x, 1/\sigma^2)$, where the j th time is in the x th stage.³

An assumption of this model is that the variability around the mean is the same for each stage, consistent with the idea that time-to-time fluctuation is a general property of different people expressing opinions at different times and does not depend on the rate of positive opinion itself, nor on the item being evaluated. Obviously, it would be simple to consider an alternative model in which there are different levels of heterogeneity in the crowd opinion for different stages, different items, or both. The specific prior chosen for σ is based on its interpretation as the period-to-period variability in the rate of up-voting that might reasonably be expected. It assumes that the standard deviation of this variability follows a positive-truncated Gaussian distribution with a mean of 0 and standard deviation of 0.05. This corresponds to expecting 0%–5% fluctuation over time periods to be common, and fluctuations up to 10% being plausible. Again, it would be simple to consider different assumptions.

The JAGS script below implements the graphical model in Fig. 7. The key changes arising from the hierarchical extension are the sampling of theta from a Gaussian defined by mu and sigma and the introduction of priors on mu[i,j] and sigma.

```
model{
  for (i in 1:m){
    for (j in 1:n){
      # response count data
      u[i, j] ~ dbin(theta[i, j], t[i, j])
      # hierarchical up-vote rate for time period
      theta[i, j] ~ dnorm(mu[i, w[i, j]], 1/sigma^2)T(0, 1)
      # stage for time period
      w[i, j] = sum(wTmp[i, j, 1:gamma]) + 1
      for (k in 1:gamma){
        wTmp[i, j, k] = step(j - tau[i, k])
      }
    }
    # prior for mean rate of stages
    for (k in 1:(gamma+1)){
      mu[i, k] ~ dunif(0, 1)
    }
    # prior for change points
    for (k in 1:gamma){
      tauPrime[i, k] ~ dcat(alpha)
    }
    # order constraint
    tau[i, 1:gamma] = sort(tauPrime[i, 1:gamma])
  }
  # spike-and-slab prior
  alpha[1] = n/(2*n - 1)
  for (k in 2:n){
    alpha[k] = 1/(2*n - 1)
  }
  # prior for standard deviation
  sigma ~ dnorm(0, 1/(0.05)^2)T(0, 1)
}
```

³ This is a truncated Gaussian distribution on the interval (0, 1), parametrized in terms of a mean and a precision.

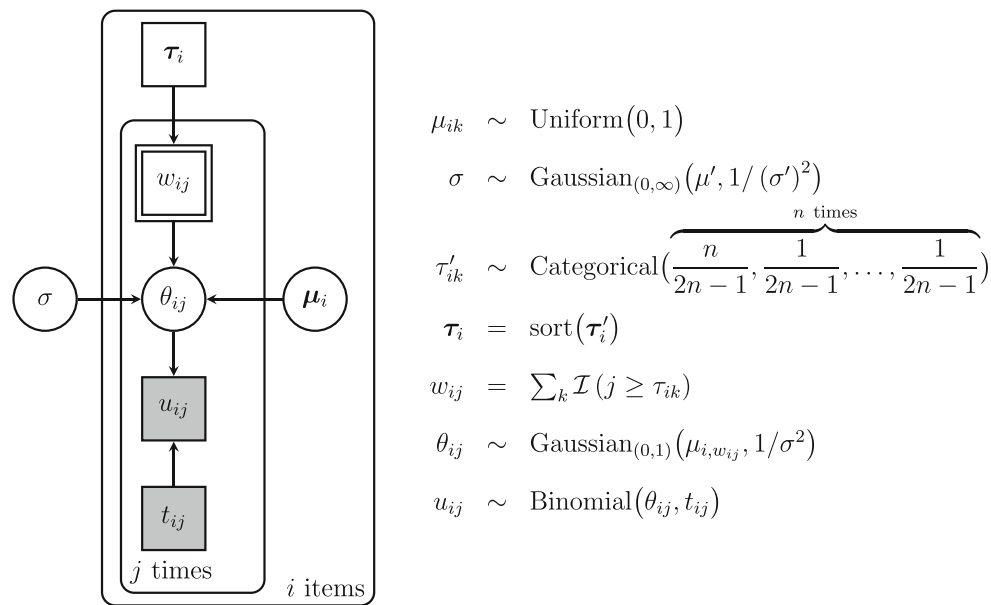


Fig. 7 Graphical model representation of the hierarchical extension of the change-point model

As before, this script is used to find the mode of the change points for each item from the joint posterior. These tau values are used as input for a second script. The second script thus makes inferences about the mean group opinion μ conditioned on the mode for tau. The second script is not shown, but it simply removes the sections that define priors for the tau parameters. It is available in the [supplementary material](#).

Modeling results

The results of applying the model to these eight players, with the setting $\gamma = 5$, are presented in Fig. 8. Each panel corresponds to a player, and the player's pattern of votes is shown by circles. The line shows the inferred mean rate of up-voting, with the dates and change points labeled. For example, in the first panel, Aaron Rodgers is inferred to have a mean rate of up-voting of about 35% until the 3rd of January, when the rate increases to about 60%. There is then a two-day drop to about 10% from the 23rd to the 25th of January, before the rate returns to around 50% until the final day. The large but brief drop in the rate has a natural interpretation as an immediate overreaction and then recalibration to Rodgers' Green Bay Packers team's elimination from the playoffs on the 22nd of January.

The inferred changes for the other players in Fig. 8 are similarly reasonable in terms of the voting data, and interpretable in terms of real-world events affecting their MVP chances. For example, the large drop in the rate of up-voting for Derek Carr occurs at about the time he suffered a season-ending injury. The two increases in the rate of up-voting for Matt Ryan occur immediately after playoff wins for his Atlanta Falcons team, in which he performed especially well as the team's

quarterback. Many of the players have an inferred change point on the 3rd of January, immediately after the regular season games were completed. We interpret these changes as the crowd reevaluating the likely MVP winner at the transition from the regular season to the playoffs.

These data provide a good example of the real-world usefulness of tracking changes in the rate of up-voting. The best prediction of the winner of the MVP award is the player with the highest rate at the time the decision is made, at the end of the voting period. If change points are detected in the rate of up-voting, the measure of *current* opinion provided by our model will differ from a standard measure of *cumulative* opinion that simply measures the overall proportion of up-votes. Formally, the current opinion is the final value of μ_i for the i th player, whereas a player's cumulative opinion at the j th time is $\sum_{k=1}^j u_{ik} / \sum_{k=1}^j t_{ik}$.

Figure 9 shows the cumulative and current opinion for all eight players. The cumulative measure, in the top panel, ranks Ryan in sixth place, behind Ezekiel Elliott, Dak Prescott, Tom Brady, Sean Lee, and Aaron Rodgers. Elliott and Prescott were both rookie players for the Dallas Cowboys and were widely viewed as outstanding new talent in a high-profile franchise that had a historically good regular season. A reasonable summary of media coverage of their MVP prospects is that early excitement faded to a more realistic assessment that rookie winners are unlikely, especially given that both played for the same team. Brady and Rodgers, in contrast, were probably the most established and high-profile perennial favorites for the MVP award. The current opinion measure, in the bottom panel, which is simply the combination of the results for the individual players shown in Fig. 8, correctly predicts Matt Ryan as the winner, on the basis of the most recent opinion. It is

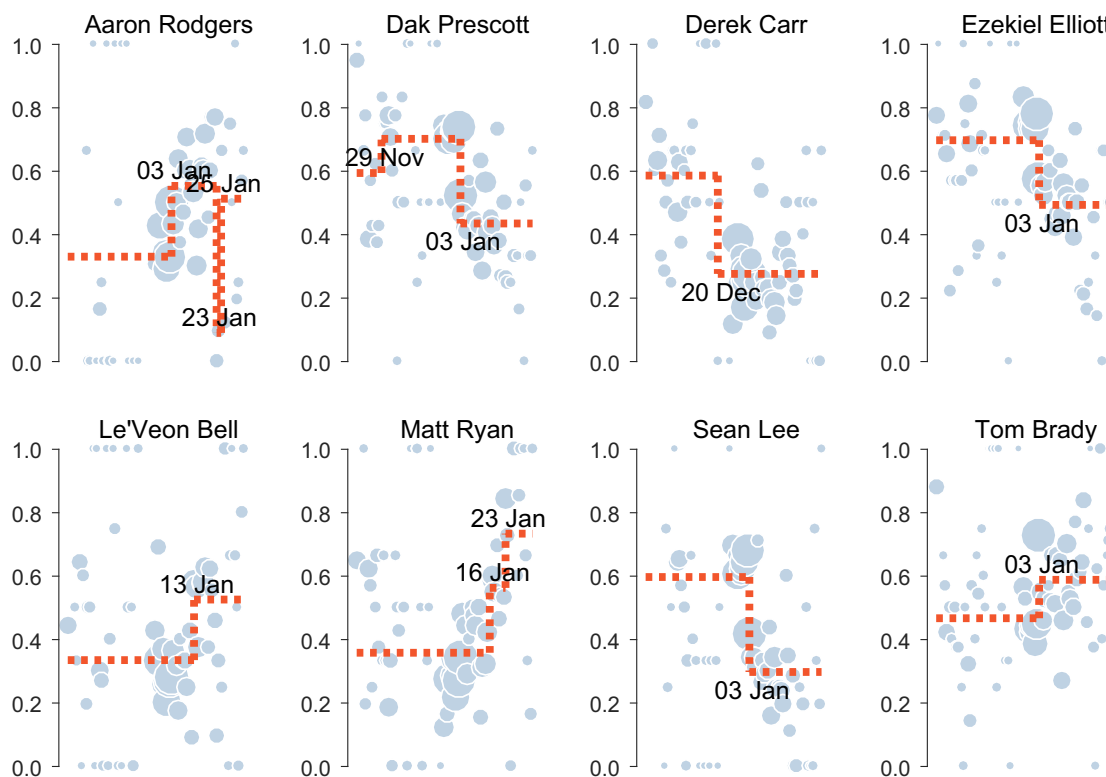


Fig. 8 Model results for crowd-sourced binary voting time series data on eight candidates for the NFL MVP award in the 2016–2017 season. Each panel corresponds to a player. The circles show the proportions of up-votes for each item on each day, and the area of each circle is proportional

to the total number of votes on that day. The broken lines show the inferred mean rates of up-voting over time, with the dates of change points labeled

interesting to note that Brady ranks second according to current opinion, and it is probably fair to speculate that he was regarded as the only other serious possible winner in the days before the award was announced.

Discussion

Sudden changes in measured behavior occur often throughout the cognitive sciences. At the individual level, people can switch their strategies, have sudden insights, or change their opinions. At the group level, different people can be assigned to tasks, or relationships between people can change quickly and drastically. Although step change is common, it presents a data analysis and modeling challenge that is not naturally handled by the default statistical methods used in the psychological sciences. Changes that are gradual—such as steady increases, smooth cycles, or random walks—are generally captured well by inherently continuous statistical models like regression, in which small changes in the inputs lead to small changes in the outputs. Modeling step change—requiring identifying where change points occur and how behavior changes as these points are crossed—presents a different statistical challenge.

We have developed and demonstrated a method for modeling step change that fits within a large statistical literature on the topic (Adams & MacKay, 2007; Barry & Hartigan, 1993; Chib, 1998; Fearnhead, 2006; Stephens, 1994). A statistical attraction of our approach is that it allows for the number of change points to be inferred together with the locations of the changes. The prior on the latent parameters corresponding to change points combines model selection with parameter estimation and provides a neat Bayesian solution to the two related inference problems of determining how many change points there are and where they are located.

The statistical approach developed here does require setting a value γ that sets the maximum number of change points and places a prior distribution on the number of change points. This part of the model could be extended in two useful ways. One possible extension would be to allow greater flexibility in setting the relative mass of the marginal spike-and-slab priors at the spike that corresponds to the absence of a change point. The present use of one-half is a sensible default, but other possibilities would allow for a broader range of prior distributions over the number of change points to be expressed. The other possible extension would be to treat γ as a dispersion hyper-parameter, much as in nonparametric Bayesian models like Chinese restaurant processes

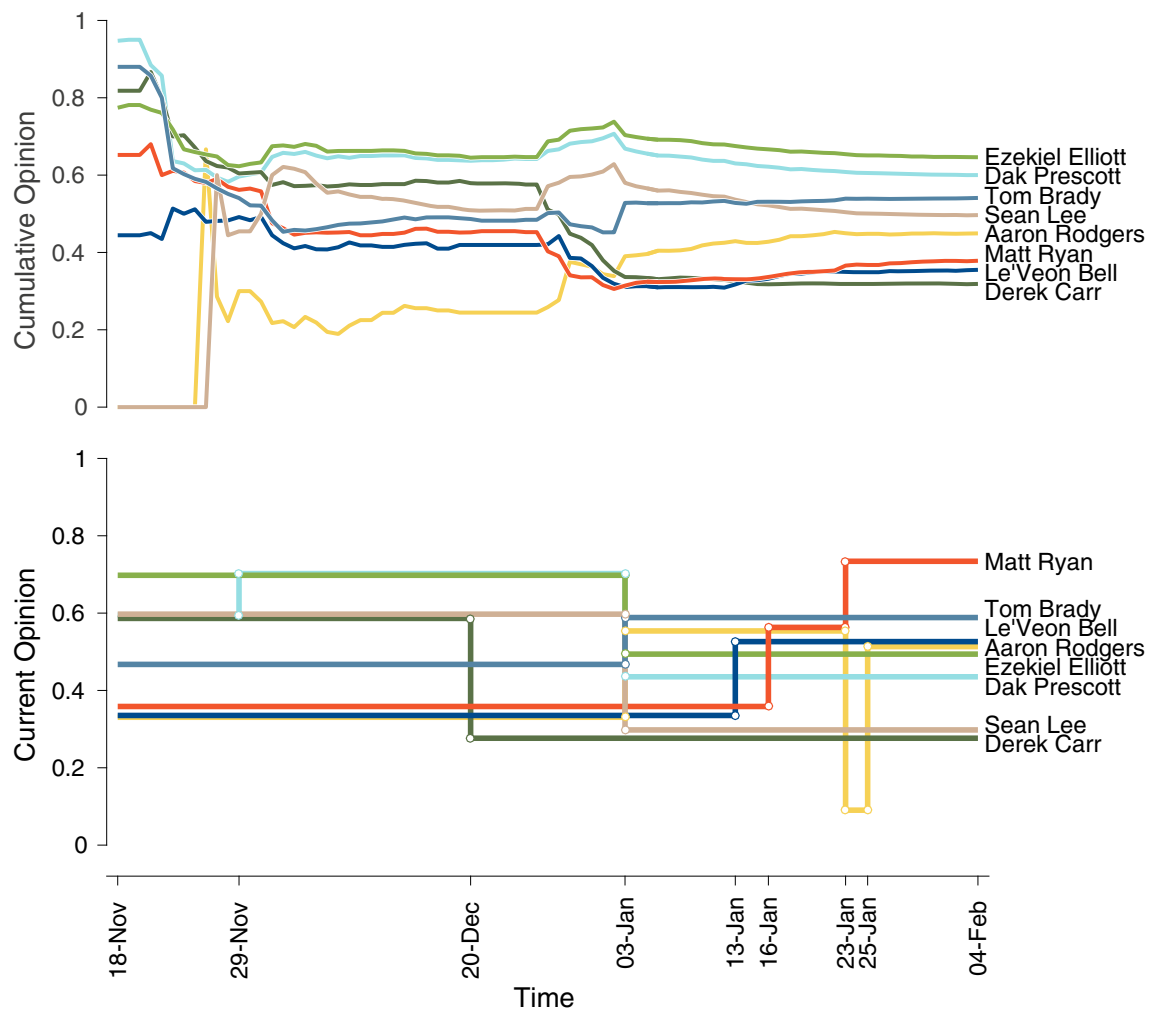


Fig. 9 Two analyses of the crowd-sourced voting data for the NFL MVP award in the 2016–2017 season. The top panel shows the cumulative opinion for eight leading candidates. The bottom panel shows the current opinion inferred by the hierarchical change-point model, based

on the inference of change points in the mean rates of up-voting. Cumulative opinion predicts Ezekiel Elliott as the MVP, whereas current opinion correctly predicts Matt Ryan

(see Navarro, Griffiths, Steyvers, & Lee, 2006), and make inferences about the number of change points in a more flexible, hierarchical setting.

Whatever exact method is used for inferring the number and location of change points, the attraction of our approach lies in its ease of use and flexibility. The core change-point model is implemented as a graphical model using JAGS and is easily extended within that formalism and software to tackle a diverse range of cognitive modeling problems that involve step change. We presented two worked examples to demonstrate this flexibility: one involving a latent-mixture extension to accommodate change that could be either gradual or sudden, and another involving a hierarchical extension that allows for homogeneity in the structure of the group data to be incorporated.

A final attraction of our approach is more conceptual or philosophical. The use of generative models here places the focus on cognitive assumptions rather than statistical methods.

The extended models of category learning and crowd opinions are built by making assumptions about how the observed behavior is generated over time. Making different or additional cognitive assumptions will lead naturally to a modified or extended model. For example, as was alluded to earlier, a category learning theorist might object to the exact assumptions in the linear-increasing account, and instead propose a different model to compete against the step-change model to explain individual behavior. Almost any model the theorist is likely to propose could be quickly and simply implemented as a graphical model within JAGS (Lee & Wagenmakers, 2013). Once the generative model is finalized, its application to data is automatically achieved using Bayesian methods. This means that there are no methodological “degrees of freedom” in the way that inferences about step change are made. Different models will, of course, lead to different results, but these differences can be understood in terms of the different cognitive assumptions being made. We think this “psychology

first” approach to data analysis is a productive one, since it places the emphasis on being explicit about the theoretical assumptions made and relegates statistical methods to their appropriate service role of specifying the mechanics of inference.

Author note I thank Irina Danileiko for help with the category learning data, Ravi Iyer for help with the [ranker.com](https://www.ranker.com) data, and Lucy Wu for help developing the NFL MVP application. **Supplementary material** for this article, including data and code, is available on the Open Science Framework project page <https://osf.io/mw3u2/>.

References

- Adams, R. P., & MacKay, D. J. (2007). *Bayesian online changepoint detection*. arXiv:0710.3742
- Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149–178. doi:<https://doi.org/10.1146/annurev.psych.56.091103.070217>
- Barry, D., & Hartigan, J. A. (1993). A Bayesian analysis for change point problems. *Journal of the American Statistical Association*, *88*, 309–319.
- Blakeley, L. (1949). The Lindisfame s/δ problem. *Studia Neophilologica*, *22*, 15–47.
- Brooks, S. P., & Gelman, A. (1997). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*, 434–455.
- Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics*, *86*, 221–241.
- Feamhead, P. (2006). Exact and efficient Bayesian inference for multiple changepoint problems. *Statistics and Computing*, *16*, 203–213.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, *82*, 711–732.
- Jordan, M. I. (2004). Graphical models. *Statistical Science*, *19*, 140–155.
- Koller, D., Friedman, N., Getoor, L., & Taskar, B. (2007). Graphical models in a nutshell. In L. Getoor & B. Taskar (Eds.), *Introduction to statistical relational learning* (pp. 13–55). Cambridge: MIT Press.
- Kruschke, J. K. (2008). Models of categorization. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 267–301). New York: Cambridge University Press.
- Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. New York: Cambridge University Press.
- Lodewyckx, T., Kim, W., Tuerlinckx, F., Kuppens, P., Lee, M. D., & Wagenmakers, E.-J. (2011). A tutorial on Bayes factor estimation with the product space method. *Journal of Mathematical Psychology*, *55*, 331–347. doi:<https://doi.org/10.1016/j.jmp.2011.06.001>
- Mitchell, T. J., & Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, *83*, 1023–1032.
- Navarro, D. J., Griffiths, T. L., Steyvers, M., & Lee, M. D. (2006). Modeling individual differences using Dirichlet processes. *Journal of Mathematical Psychology*, *50*, 101–122. doi:<https://doi.org/10.1016/j.jmp.2005.11.006>
- Pettitt, A. (1979). A non-parametric approach to the change-point problem. *Journal of the Royal Statistical Society: Series C*, *28*, 126–135.
- Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In K. Hornik, F. Leisch, & A. Zeileis (Eds.), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing* (pp. 1–10). Vienna: R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf>
- Ross, A. S. C. (1950). Philological probability problems. *Journal of the Royal Statistical Society: Series B*, *12*, 19–59.
- Rouder, J. N., Haaf, J. M., & Vandekerckhove, J. (2018). Bayesian inference for psychology, part IV: Parameter estimation and Bayes factors. *Psychonomic Bulletin & Review*, *25*, 102–113. doi:<https://doi.org/10.3758/s13423-017-1420-7>
- Silvey, S. (1958). The Lindisfame scribes’ problem. *Journal of the Royal Statistical Society: Series B*, *20*, 93–101.
- Smith, A. F. M. (1975). A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika*, *62*, 407–416.
- Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1411–1436. doi:<https://doi.org/10.1037/0278-7393.24.6.1411>
- Stephens, D. A. (1994). Bayesian retrospective multiple-changepoint identification. *Journal of the Royal Statistical Society: Series C*, *43*, 159–178. doi:<https://doi.org/10.2307/2986119>
- Venter, J., & Steel, S. (1996). Finding multiple abrupt change points. *Computational Statistics & Data Analysis*, *22*, 481–504.