

A General Latent Assignment Approach for Modeling Psychological Contaminants

Matthew D. Zeigenfuse Michael D. Lee

Department of Cognitive Sciences
University of California, Irvine

Abstract

Data from psychological experiments are rife with ‘contaminants’, which can generally be defined as data generated by psychological processes different from those intended as the object of study. Contaminant data can interfere with the testing of substantive psychological models and their parameters, so it is important to have methods for their identification and removal. After noting that current practices in cognitive modeling for dealing with contaminants are not completely satisfactory, we argue for a general latent mixture approach to the problem. We demonstrate the tractability and effectiveness of the approach concretely, through a series of four applications. These applications involve a simple choice problem, a diffusion model of a response time and accuracy in decision-making, a hierarchical signal detection model of recognition memory, and a reinforcement learning model of decision-making on bandit problems. We conclude that developing models of contaminant processes requires the same sort of creative effort that is needed to model substantive psychological processes, but that it is a necessary endeavour that can be coherently and usefully pursued within the latent mixture modeling approach.

Introduction

In most psychological experiments, at least some participants will complete the task in a way that is inconsistent with the motivating theoretical and modeling goals of the study. Diligent participants can fail to understand the task, and less diligent participants can find quick and effortless ways to complete the task that do not involve interesting cognitive processes. For these, and a host of other reasons, psychological data are usually rife with ‘contaminants’, which can generally be defined as data generated by psychological processes different from those intended as the object of study. It is widely recognized in both the statistical and model-based analysis of

psychological data that contamination is an important practical issue. In particular, it is understood that contaminant data can change or blur what is learned from the remainder of a data set.

For such a basic and ubiquitous problem, however, the development and use of theoretical methods for identifying and removing contaminants receives surprisingly little direct attention. We think that current practices are both inconsistent and *ad hoc*, and suspect that often contaminants are removed without explicit mention in research paper. Looking at recent cognitive modeling research—where one might expect some level of sophistication, given the impressive quantitative skills evident in the research—it seems clear that, when contaminant removal methods are made explicit, different authors adopt different methods, and sometimes the same authors adopt different methods in different papers. Very often, the removal of contaminant participants is based on simple statistical summaries of the observed data (e.g., Little & Lewandowsky, 2009; Little, Lewandowsky, & Heit, 2006; Zeithamova & Maddox, 2006), pre-determined performance criteria (e.g., Goldstone, Lippa, & Shiffrin, 2001), or behavior consistent with not understanding the task (e.g., Zacks, 2004). Sometimes these methods are presented with a disclaimer (e.g. Rouder, Sun, Speckman, Lu, & Zhou, 2003) that a more sophisticated modeling approach would have been theoretically preferable. We agree that, while none of these methods are unreasonable, and most probably suffice for their practical goal of tidying the data, better and more principled modeling approaches are possible and desirable.

This paper argues for a general latent assignment approach to dealing with contaminants in psychological data. This is a well known and understood approach in statistics (e.g., Tanner & Wong, 1987), and its straightforward adaptation to the problem of contaminants is also well documented (e.g., Jaynes, 2003, Ch. 21). The framework operates by assuming that each observed datum was generated either by a substantive model of the cognitive processes being studied, or by one or more contaminant models. Treating these substantive versus contaminant assignments as non-ignorable missing data gives a likelihood function that, when combined with appropriate prior distributions over model parameters, allows inferences to be made about both the substantive model parameters, and about which data are contaminants.

We demonstrate these general and useful features through a series of four cognitive modeling application. The first involves a simple model of choice, and provides a toy example to make the basic ideas concrete, including the ability of contaminants to change inferences about psychological models. The second example involves Wiener diffusion model of two-alternative forced-choice decision-making, and illustrates how contaminants can change the estimation of key model parameters. The third example involves a signal detection model of recognition memory, showing how the approach extends to hierarchical models, and beginning to address the issue of how the identification of contaminants depends on psychological theory. The fourth example expands on this important issue, using heuristic models of decision-making on bandit problems

to show that the choice of contaminant model or models can significantly affect inference for parameters of the substantive model of interest. We conclude by discussing the challenge of developing contaminant models—which we regard as involving the same difficult creative exercise as developing models of substantive psychological processes themselves—while arguing that the latent mixture modeling approach provides a general and coherent framework for their incorporation into modeling.

Currently Used Methods for Contaminants

Methodologies for dealing with contaminant data are not new (e.g., Barnett, 1978; Box, 1979; Rudas, Clogg, & Lindsay, 1994). The approach we advocate dovetails nicely with the two dominant approaches applied in psychological literature. The simplest, and by far most prevalent, approach, is trimming (e.g. Kutner, Nachtsheim, Neter, & Li, 2004). This approach rank orders the observed data and removes either some percentage, typically 1 – 5%, of the most extreme points from consideration, or those points exceeding some cut-off, such as three standard deviations from the mean. This method can be thought of as providing a model-free point estimate of the correct configuration in the space of all possible substantive versus contaminant assignments.

While often appropriate with a large number of data, trimming is more problematic with only a few. Trimming corresponds to the *ad hoc* assumption that the data to be trimmed are in the tails of the likelihood. This may often be reasonable when we have a large number of observations, but it can be unreasonable for small numbers of data, since the probability of seeing any data in the tails is low. And it can also be inappropriate when likelihoods are highly non-normal, which can be the case for complicated cognitive models. In general, of course, data are valuable and we would like to throw away as few as possible.

In recognition of these problems, some cognitive modeling uses a more sophisticated mixture approach (e.g., Ratcliff & Tuerlinckx, 2002). This approach assumes each observed datum is generated by a process whose likelihood is a mixture of substantive and contaminant likelihoods. This approach rectifies the potential inconsistency of trimming the raw data, and thus produces appropriate parameter estimates for the substantive model.

Despite these strengths, however, the mixture likelihood approach can be thought of as being at “too high a level” for many psychological applications, because it fails to infer some key information. In particular, it does not identify explicitly which data are the contaminants. Ideally, we would like a model which combines the parameter inference of mixture likelihoods with a more principled approach to identifying the contaminant data than that provided by trimming.

The Latent Assignment Approach

Latent assignment provides such an approach by rejoining to the mixture likelihood approach the ability to specify substantive and contaminant data. Formally, it does that by expanding the mixture likelihood in terms of the latent specific versus contaminant assignments. That is, in the latent assignment approach, each datum is treated as intrinsically either substantive or contaminant. Theoretically, therefore, latent assignment is naturally conceived as data augmentation, which is a familiar approach from statistics, and can be traced back to the seminal paper by Dempster, Laird, and Rubin (1977). In practice, the latent assignment approach to contaminant modeling offers the best of both worlds: it offers the consistent parameter estimates of mixture likelihoods and the contaminant identification provided by trimming.

For a particular experiment, suppose we observe N data points, designated $\mathbf{Y} = (y_i)_i$, such that each y_i believed to have been generated by one of two stochastic models, a substantive model p^S parameterized by θ_S and a contaminant model p^C parameterized by θ_C . Furthermore, define missing data $\mathbf{Z} = (z_i)_i$ as follows: $z_i = 1$ if i th data point was generated by the substantive model and $z_i = 0$ if it was generated by the contaminant model. Finally, introduce a parameter ϕ , which can be thought of as the rate at which participants generate substantive data. We relate these parameters via the model

$$\begin{aligned} y_i &\stackrel{iid}{\sim} \begin{cases} p^S(\cdot|\theta_S), & \text{if } z_i = 1 \\ p^C(\cdot|\theta_C), & \text{if } z_i = 0 \end{cases} \\ z_i &\stackrel{iid}{\sim} \text{Bernoulli}(\phi), \end{aligned} \tag{1}$$

which has likelihood

$$L(\theta_S, \theta_C, \phi | \mathbf{Y}, \mathbf{Z}) = \prod_{i=1}^N \left\{ \phi^{z_i} (1 - \phi)^{1-z_i} \left[z_i p^S(y_i | \theta_S) + (1 - z_i) p^C(y_i | \theta_C) \right] \right\}. \tag{2}$$

It should furthermore be noted that given \mathbf{Y} and \mathbf{Z} , θ_S , θ_C , and ϕ are independent, as their likelihood factors into

$$\left[\phi^{\sum z_i} (1 - \phi)^{N - \sum z_i} \right] \left[\prod_{\{i: z_i=1\}} p^S(y_i | \theta_S) \right] \left[\prod_{\{i: z_i=0\}} p^C(y_i | \theta_C) \right]. \tag{3}$$

If we observed a particular configuration \mathbf{Z} , we could easily perform inference via Bayes' rule or maximum likelihood estimation using Equation (2); however, as much as one may wish they would, participants rarely report whether or not they employed the substantive mental processes in question on a given task, or trial within a task. Thus, we can imagine the missing data problem as having been created by a less-than-competent research assistant who forgot to transcribe some of the data from our experiment.

Of the many possible methods for dealing with this problem, we are focusing on mixture likelihoods and latent assignment. Mixture models get their name from

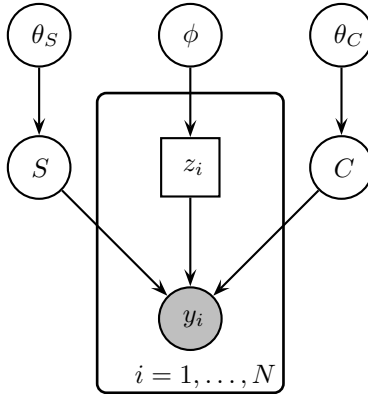


Figure 1. Graphical model illustrating missing data approach.

the fact that their likelihoods are weighted mixtures of other likelihoods. In the case at hand, the mixture comes from marginalizing over the z_i to form a new likelihood,

$$L(\theta_S, \theta_C, \phi | \mathbf{Y}) = \prod_{i=1}^N \left[\phi p^S(y_i | \theta_S) + (1 - \phi) p^C(y_i | \theta_C) \right]. \quad (4)$$

Inference can be performed using maximum likelihood estimation, Bayesian updating, or any other compatible scheme using this missing data-free likelihood.

Marginalizing, however, is not the only way to deal with the missing data; one could forgo forming the mixture likelihood by assigning prior distributions over the parameters and performing inference via Bayesian updating. The structure of resulting model is shown in the graphical model in Figure 1. Graphical models (for statistical and psychological introductions see, respectively Jordan, 2004; Lee, 2008) are a means for representing the conditional dependence structure of a given statistical model using directed, acyclic graphs¹.

Graphical models illustrate dependence genealogically, with arrows pointing from parent to child: parents are dependent only through their children. Hence, upon conditioning on these parents, they become independent. For our basic model in Equation (1), the graphical model shows that θ_S , θ_C , and missing data are random variables, the former two continuous (denoted by circular nodes), the latter discrete (denoted by square), which are independent given the observed (denoted by shaded) data, and moreover that, upon observing the missing data, ϕ and θ_S and θ_C also become independent.

¹Directed, acyclic graphs are not the only type of graphical model. Inference methods have also been developed for undirected, or associative, graphs (see Pearl, 2000, for details).

Example 1: A Simple Choice Problem

The goal of our first example is to present a very simple modeling problem that makes clear the general problem potentially posed by contaminants, and shows how the latent assignment approach can remedy these problems. In particular, this example emphasizes the point that *contaminants can lead to the wrong generating model for data being inferred*.

Background

Consider a simple two-alternative choice experiment, in which the research question is whether people prefer one alternative to the other. One way to test this hypothesis uses a parameter-free model positing that the observed data \mathbf{k} follow unbiased choosing, and so are independent and binomially distributed with the common rate $\theta = 1/2$. Suppose the experiment consists of $m = 100$ trials, and the observed data are $\mathbf{k} = (50, 50, 55, 45, 50, 55, 45, 50, 55, 45, 50, 55, 50, 52, 55, 99)^T$ for $N = 16$ participants. With the exception of the last participant, these data are consistent with the unbiased model. When, however, we compare this model $H_0 : \theta = 1/2$, to a more complex, parameterized model with a single free parameter, representing the bias with which participants choose the first alternative, $H_1 : 0 < \theta < 1$, the Bayes factor,

$$\text{BF} = \frac{p(D \mid H_1)}{p(D \mid H_0)},$$

comes out to be approximately 3.3, in favor of the more complicated model showing bias in choice.

This simple example illustrates the danger of performing inference without taking contaminant data into account. Indeed, the data observed for the last participant could very plausibly been generated by a very common type of contaminant, a participant giving the same response on almost every trial without even attempting to do the task. Though, for this example, trimming gives the same result as the more complicated latent assignment approach, we use it throughout in order to illustrate how the formal latent assignment framework can be applied to particular problems.

Model

The latent assignment approach to contaminants uses two models for data generation: a substantive model, modeling the cognitive processes we believe generated the observed data, and a contaminant model, modeling how contaminants are generated. In this case our substantive model specifies that the observed data were generated by a binomial likelihood with rate parameter $\theta = 1/2$. For the contaminant model, this simple example does not allow psychological theory to define the contaminant model. Instead, we rely on simple statistical considerations. If we know that both choices are possible in the task, but know nothing more about contamination, then there is a

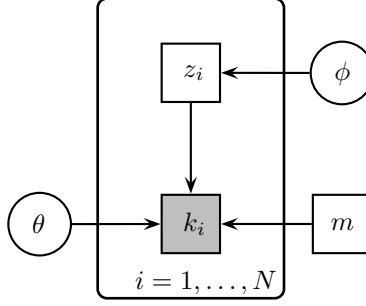


Figure 2. Graphical model for the rate problem.

compelling argument, based on invariance properties from objective Bayesian analysis, that the uniform prior $\theta \sim \text{Uniform}(0, 1)$ is the unique correct prior distribution (see Jaynes, 2003, pp. 382–385).

Placing these substantive and contaminant models into Equation (1), gives the latent assignment model

$$\begin{aligned}
 k_i | z_i &\sim \begin{cases} \text{Binomial}(m, \frac{1}{2}), & \text{if } z_i = 1 \\ \text{Binomial}(m, \theta), & \text{if } z_i = 0 \end{cases} \\
 z_i | \phi &\stackrel{iid}{\sim} \text{Bernoulli}(\phi) \\
 \phi &\sim \text{Uniform}(0, 1).
 \end{aligned}$$

where we have used the same result regarding uniform priors on rates to justify the prior on the contaminant rate ϕ .

The corresponding graphical model is shown in Figure 2. This figure shows the conditional dependencies between \mathbf{k} , \mathbf{z} , ϕ and θ . The arrows pointing from z_i and θ to k_i illustrate that the k_i are independent having conditioned on z_i and θ , since m is fixed. The arrow from ϕ to z_i illustrates that the z_i are conditionally independent given ϕ .

Analysis

The graphical model in Figure 2 was implemented in WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000; Lunn, Spiegelhalter, Thomas, & Best, in press), which uses Markov Chain Monte Carlo (MCMC) to draw samples from the joint posterior distributions of θ , ϕ , and \mathbf{z} (e.g., Gilks, Richardson, & Spiegelhalter, 1996). The inferences reported here are based on four chains, seeded with random initial values, collected 60,000 samples are a burn-in of 10,000 discarded samples. We used the standard \hat{R} statistic, comparing within-chain variance and between-chain variance (Gelman, Carlin, Stern, & Rubin, 2004), as a diagnostic measure to assess and ensure convergence.

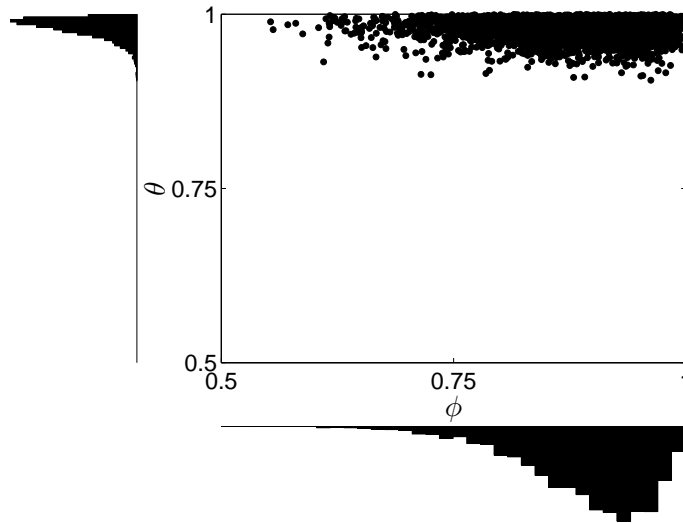


Figure 3. Joint posterior over ϕ , the contaminant rate, and θ , the biased choice rate parameter. Histograms of samples from marginal posterior distributions over each variable lie along their respective axes. A scatterplot these pairs is shown in the center plot.

The advantage of latent assignment over expected likelihood models is the ability to infer a probability distribution over potential configurations of the data. For our choice problem, this space has 2^{16} elements, which is a small space of potential configurations in the context of a typical psychological experiment. By construction, the mixture modeling is unable to update this probability given data, but the latent assignment model can. In this example, all of the posterior mass is given to one assignment, $z_1, \dots, z_{15} = 1$ and $z_{16} = 0$, which we label $\hat{\mathbf{z}}$.

The posterior joint probability distribution over ϕ and θ for the latent assignment model is shown in Figure 3. Inspection suggests the two variables are independent, a fact necessitated by the point mass distribution on \mathbf{Z} . To see why note that, for our purposes, sampling a point mass distribution located at $\hat{\mathbf{z}}$ is analogous to observing $\hat{\mathbf{z}}$, allowing us to condition on it. Factoring our likelihood according to Equation (3) given our flat prior distributions over θ and ϕ we get

$$p(\theta, \phi \mid \mathbf{k}, \mathbf{Z} = \hat{\mathbf{z}}) \propto \left[\phi^{\sum \hat{z}_i} (1 - \phi)^{N - \sum \hat{z}_i} \right] \left[\theta^{k_i} (1 - \theta)^{m - k_i} \right].$$

This fortuitously observed point mass distribution on \mathbf{z} allows us to derive analytic marginal posterior distributions on θ and ϕ . Integrating each parameter over its support yields

$$\begin{aligned} p(\theta \mid \mathbf{k}, \mathbf{Z} = \hat{\mathbf{z}}) &= \text{Beta}(\sum[(1 - \hat{z}_i)k_i], \sum[(1 - \hat{z}_i)(m - k_i)]) \\ &= \text{Beta}(k_{16}, m - k_{16}) \end{aligned} \tag{5}$$

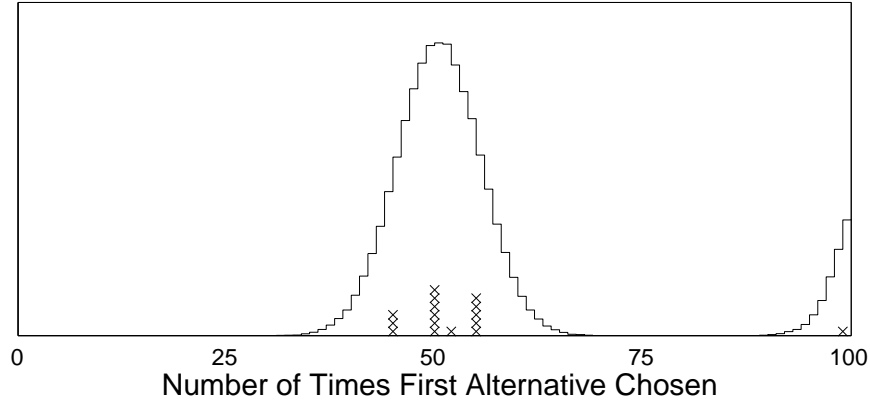


Figure 4. Posterior predictive and data for the number of times the first alternative is chosen.

for θ , and

$$p(\phi \mid \mathbf{k}, \mathbf{Z} = \hat{\mathbf{z}}) = \text{Beta}(\sum \hat{z}_i, N - \sum \hat{z}_i) \quad (6)$$

for ϕ . These distributions emphasize important aspects of the nature of contaminant inference. The form of the marginal posterior over θ underscores that inference updates the distribution over contaminant model parameters based solely on the data inferred to be contaminant. Had the substantive distribution parameters, Equation (5) would extend analogously to the substantive case with $1 - \hat{z}_i$ being replaced with \hat{z}_i inside the summations. With respect to the rate at which substantive data is generated by participants, we see ϕ is updated as a simple beta-binomial model after observing $\hat{\mathbf{z}}$.

In general cognitive modeling contexts, we will not observe data so strongly identifying our missing data. However, the previous discussion is useful for illustrating a generally applicable method of thinking about inference. Suppose inference yields posterior distribution $p(\mathbf{Z} \mid \mathbf{k})$ over the missing data. For each \mathbf{z} in the support of $p(\mathbf{Z} \mid \mathbf{k})$ on $\{0, 1\}^N$ we can update the substantive and contaminant parameters by updating given the data the configuration \mathbf{z} determines to be substantive and contaminant, respectively. Simultaneously, we can update ϕ via the beta-binomial model in Equation (6) for the same configuration. Finally, we can compute the marginal posterior over θ and ϕ by summing over these conditional posteriors weighted by the marginal posterior mass of their corresponding configuration. This illustrates the function of the missing contaminant data: to ‘smear’ inference on the parameters, hedging our bets on parameter values by our uncertainty over the correct contaminant configuration.

The posterior predictive distribution, together with the observed data, for the number of first alternative choices is shown in Figure 4. We interpret this distribution for a new participant whose data we have yet to observe, enumerating the predicted

probabilities of observing each possible number of times they might chose the first alternative. In this case, we see that much of the predictive mass is centered around $k = 50$, the prediction of our parameter-free substantive model of unbiased choosing, but also that there exists a non-negligible mass for values between 90 and 100, the posterior prediction of our contaminant model. In this way, posterior predictive distributions over data provide a powerful of method for validating a model, since predicted fit to the observed data increases as models become more correctly specified. In addition, the posterior predictive in Figure 4 also illustrates when predictive distributions over observed data may be inappropriate in contaminant modeling. Often in this type of modeling, the two models, substantive and contaminant, are of differential psychological value, the contaminant model possessing no psychological value in and of itself, only in its ability to filter contaminant points out of our inferences about the substantive model.

In this case, using a contaminant model keeps the data that intuitively seemed to characterize participants' choices, corresponding to those consistent with unbiased choice. Maintaining these data, and removing the single contaminant participant identified by the latent assignment method, the model comparison between the unbiased and biased hypotheses is now 25.5 in favor of the unbiased account.

Example 2: A Diffusion Model of Decision-Making

Our second example deals with a more complicated psychological model, involving a sequential-sampling model of accuracy and response time in decision-making. The main point demonstrated by this example is that contaminant data—indeed, a single contaminant datum—can *significantly affect inferences about key psychological parameters* for an otherwise excellent model of the data.²

Background

Sequential sampling models, in their various forms, represent a widely-used and successful approach to modeling accuracy and the distribution of response times in human decision-making (e.g. Busemeyer & Townsend, 1993; Link & Heath, 1975; Nosofsky & Palmeri, 1997; Ratcliff & McKoon, 2008; Smith, 2000; Vickers, 1979). These models assume that people sample evidence from a noisy signal generated by the stimuli and make a decision once sufficient evidence has been gathered to favor one or another alternative. When the sampling rate is assumed to be sufficiently high, sequential sampling processes can be approximated by continuous-time, stochastic diffusion processes. If it is further assumed that the evidence is sampled from a stationary Gaussian distribution, this diffusion is known as a Wiener process.

²Since the research for this example was completed, a much more complete and impressive use of latent assignment mixture modeling, combining the full Ratcliff diffusion model (Ratcliff & McKoon, 2008) with two contaminant distributions, has been reported by Vandekerckhove, Tuerlinckx, and Lee (2008). We maintain this example, however, because it makes the point about parameter estimates being affected by contaminants in a simpler but still interesting cognitive modeling context.

In its simplest form, a Wiener diffusion model of a two-alternative forced-choice task has four parameters: the drift rate v , which measures measure of the evidence provided by the stimulus; the boundary separation a , which measures the level of evidence needed to make a decision; the starting point ξ , which measures the bias in favor of one or the another decision; and the non-decision time T^{er} , which measures the time necessary for tasks such as encoding and responding. We denote the likelihood of a Wiener diffusion model having these parameters as $\text{Wiener}(v, a, \xi, T^{\text{er}})^3$.

As pointed out by Ratcliff and Tuerlinckx (2002), contaminants can severely impact diffusion parameter estimation. Examples of contaminants include a participant not encoding the evidence offered by a stimulus and responding randomly or on the basis of their bias, and a participant not paying attention and encoding evidence slowly, resulting in a large response time. In the first case, since decision time must be greater than T^{er} , this fast response will produce inferences about T^{er} which are underestimates, leaving v , ξ , and a to compensate. Long response times, on the other hand, will produce underestimates of v and overestimates of a .

Were these contaminants confined to a small subset of high leverage points, determining contaminants would be clear cut and a suitable trimming method could be found. Such straightforward situations are, however, not the norm. To deal with difficulty, Ratcliff and Tuerlinckx (2002) developed a mixture model consisting of diffusion and uniform likelihoods. In our missing data formulation, this corresponds to forming the mixture likelihood in Equation (4) with the substantive model being the diffusion model and the contaminant model being the uniform. We develop here the extended approach using latent assignment with the Wiener diffusion model⁴ and use this model to analyze part of the seminal data reported by Ratcliff and Rouder (2000).

Model

Let d_i be the the i th decision made by a participant and t_i be the response time for that decision. For a substantive trial, we consider the accuracy and response time data to be generated by a symmetric Wiener distribution whose starting point is equidistant from either boundary (i.e., $\xi = a/2$). On contaminant trials, we consider the accuracy to be generated by a Bernoulli distribution with rate $1/2$, and the response times to be drawn from a uniform distribution on $(0, \lambda)$, where λ is a parameter to be inferred (though it is constrained to be greater than the largest observed response time). As in the choice problem, we define $\mathbf{Z} = (z_i)$ to be an indicator variable whose value is 0 if the i th trial is a contaminant and 1 if it is not. The base-rate variable ϕ continues to represent the proportion of non-contaminants in the observed data.

³We omit functional form because it does not impact our discussion. The interested reader can refer to Vandekerckhove et al. (2008).

⁴Actually, Ratcliff and Tuerlinckx (2002) use the full Ratcliff diffusion model, not the simpler Wiener model, but the contaminant modeling principles applied are the same in either case.

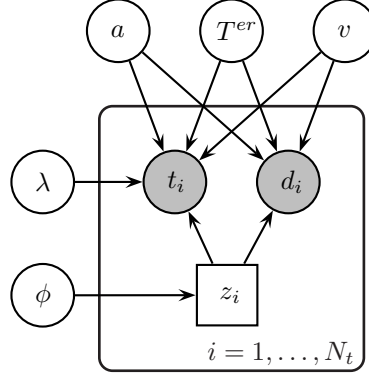


Figure 5. Graphical model for the Wiener diffusion analysis using the latent assignment contaminant framework.

These definitions yield the graphical model shown in Figure 5. We derive the formal model in Equation by placing prior distributions over the remaining model parameters. A standard normal distribution is placed on v . Since we believe *a priori* that a is both positive and contained in a somewhat small neighborhood about 1, we use a $\text{Gamma}(1, 1)$ distribution. For convenience we define T^{er} to be uniform on $(0, 2)$. However, as shown in the parameter estimates below, this interval is wide enough not to effect significantly inferences about the Wiener parameters. Since we want non-zero probability that a given observed response time is contaminant, but at the same time want the data to determine the appropriate width of the response time contaminant distribution, we define λ to be $\text{Gamma}(\epsilon, \epsilon)$ on $\lambda \in (\max_i(t_i), \infty)$ and 0 otherwise. As in the rate problem, we define ϕ to be uniformly distributed on $(0, 1)$. Thus, the generative model for the data is

$$\begin{aligned}
 d_i, t_i \mid z_i, v, a, T^{\text{er}}, \lambda &\sim \begin{cases} \text{Bernoulli}(0.5), \text{Uniform}(0, \lambda), & \text{if } z_i = 0 \\ \text{Wiener}(v, a, a/2, T^{\text{er}}), & \text{if } z_i = 1 \end{cases} \\
 z_i \mid \phi &\sim \text{Bernoulli}(\phi) \\
 a &\sim \text{Gamma}(1, 1) \\
 T^{\text{er}} &\sim \text{Uniform}(0, 2) \\
 v &\sim \text{Gaussian}(0, 1) \\
 \lambda &\sim \text{Gamma}(\epsilon, \epsilon) \times I_{\{\lambda \in \mathbb{R}: \lambda > \max_t(t)\}} \\
 \phi &\sim \text{Uniform}(0, 1).
 \end{aligned}$$

Analysis

We applied this latent assignment model to analyze participant KR out of the three participants reported by Ratcliff and Rouder (2000). The posterior distribution

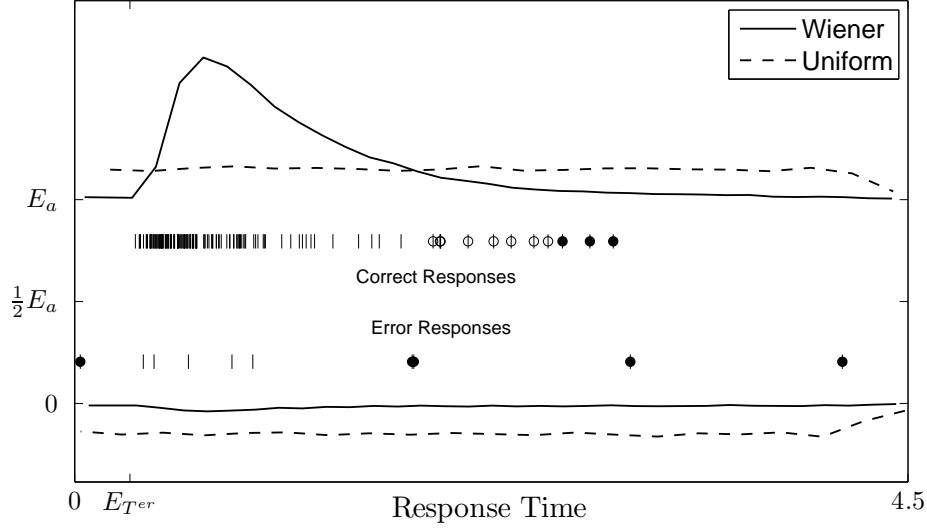


Figure 6. Posterior predictive Wiener distribution with uniform contaminant distribution. The solid line corresponds to posterior predictive Wiener distribution, the dashed to the uniform. E_{Ter} stands for $E[T^{er} \mid \mathbf{d}, \mathbf{t}]$ and E_a for $E[a \mid \mathbf{d}, \mathbf{t}]$. The observed data are shown between curves with hash marks. Hash marks without circles have $E[z_i \mid \mathbf{d}, \mathbf{t}] > 0.5$, hash marks with open circles have $0.1 < E[z_i \mid \mathbf{d}, \mathbf{t}] < 0.5$, and hash marks with closed circles have $E[z_i \mid \mathbf{d}, \mathbf{t}] < 0.1$.

was again sampled numerically using MCMC, using four chains with random initial values, and 5,000 collected samples after a burn-in of 100,000 samples.

Posterior over Assignments. The results of this analysis are presented in Figure 6, which shows the posterior predictive distributions over response times for both the substantive and contaminant models, plotted alongside the observed data. Points and curves above $\frac{1}{2}E[a \mid \mathbf{d}, \mathbf{t}]$ correspond respectively to the observed response times and posterior predictive densities for correct responses, while those below correspond to the same quantities for error responses. For the i th datum, the proportion of posterior samples in which $z_i = 1$ approximates computationally $E[z_i \mid \mathbf{d}, \mathbf{t}]$, which is the inferred posterior probability that the decision was substantive. In Figure 6, observed decisions are represented by hash marks, those marks without circles have $E[z_i \mid \mathbf{d}, \mathbf{t}] > 0.5$, those with open circles have $0.1 < E[z_i \mid \mathbf{d}, \mathbf{t}] < 0.5$, and those with closed circles have $E[z_i \mid \mathbf{d}, \mathbf{t}] < 0.1$.

Figure 6 makes clear that the posterior Wiener density of an observed decision relates to the probability the point is substantive. It shows points in regions of high density for correct and error response time distributions are assigned to Wiener over 50% of posterior samples, points in regions of small, but non-negligible Wiener density are assigned to Wiener in between 10 and 50% of samples, and points with

negligible Wiener mass are assigned to the contaminant distribution in over 90% of samples. This illustrates both the ability of the latent assignment to capture the degree to which each data point is fit by each model and the high degree with which the Wiener model captures the distribution of the observed data.

Effect of ϕ on Assignments. On closer inspection, the intuitive substantive assignment probability estimates reveal an interesting aspect of latent assignment modeling. In the error distribution, even though the points assigned lie in the higher probability region of the error response time distribution, the contaminant distribution has more mass. Intuitively, this occurs because the Wiener model is able to fit the correct response times so well, causing a large number of points to be inferred to the substantive distribution, which in turn lead the model to sample only large values of ϕ , as shown in the top panel of Figure 7.

Intuitively, the z_i are dependent in a manner analogous to the way in which flips of a coin are dependent if we do not know whether it is biased. Suppose we have seen 100 flips of a coin and are asked to predict the 101st. If we have seen 50 heads and 50 tails in the first 100 we may justifiably have no preference toward one or the other alternative, whereas if we have seen 100 heads and no tails we would be unwise to choose tails. In the current situation, latent assignment does much the same: clearly substantive points are assigned to the substantive model with high probability; clearly contaminant points are assigned to the contaminant model; and the remaining points are inferred with bias toward the substantive model commensurate with the marginal posterior shown in the top panel of Figure 7.

We can make this more formal by turning once again to the factored likelihood in Equation (3). Fix θ_S , θ_C , and ϕ and let \mathbf{z} be a particular contaminant configuration. Suppose we move the configuration \mathbf{z} to a new configuration \mathbf{z}' by flipping z_i to $1 - z_i$ for a single i and holding z_j fixed for all $j \neq i$. Further assume without loss of generality that $z_i = 0$ and $z'_i = 1$. Consider the ratio between the likelihoods of the two points

$$\frac{L(\mathbf{z} \mid \theta_S, \theta_C, \phi, \mathbf{Y})}{L(\mathbf{z}' \mid \theta_S, \theta_C, \phi, \mathbf{Y})} = \frac{(1 - \phi) p^C(y_i \mid \theta_C)}{\phi p^S(y_i \mid \theta_S)}.$$

Then, $L(\mathbf{z} \mid \theta_S, \theta_C, \phi, \mathbf{Y}) > L(\mathbf{z}' \mid \theta_S, \theta_C, \phi, \mathbf{Y})$, i.e. the likelihood the i th data point is contaminant is greater than the likelihood it is substantive, all else equal, implies

$$\frac{p^C(y_i \mid \theta_C)}{p^S(y_i \mid \theta_S)} > \frac{\phi}{1 - \phi}. \quad (7)$$

Thus, given a particular values of θ_S , θ_C , and ϕ and holding all other substantive versus contaminant assignments constant, the posterior probability of that the i th datum is contaminant is greater than the posterior probability it is substantive only when Equation (7) holds.

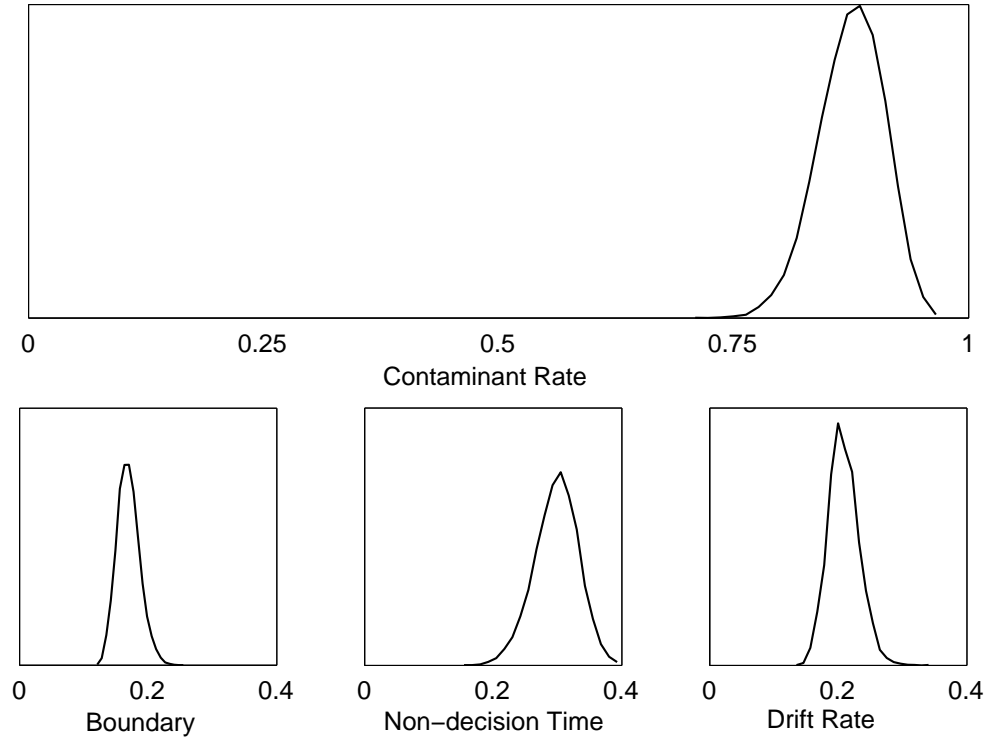


Figure 7. Posteriors over the non-contaminant rate ϕ , and the Wiener model boundary separation a , non-decision time T_{er} and drift rate v parameters.

Parameter Inference. Posterior distributions over the three model parameters, v , a , and T^{er} are shown in the bottom panels of Figure 7. These distributions illustrate the way in which the contaminant inference influences inference about parameter values. As discussed in previous sections, for a given substantive versus contaminant configuration inference on the parameters of the substantive model depends only upon those points determined to be substantive. Thus, the influence of a given point on parameter inference is commensurate with its marginal posterior mass. This is most evident in the marginal posterior over T^{er} , which gives non-negligible density only to those points within roughly the open interval $(0.2, 0.4)$. The smallest observed response time, however, is 0.030s, which lies well outside this interval. Figure 6 shows that the marginal posterior mass for this point is less than 0.1 (in fact, it is extremely close to 0). Through this clear latent assignment as a contaminant response time, the 0.030 datum does not affect the estimation of the key T^{er} parameter.

One point we want to emphasize is that the contaminant model, given equal probability to both decisions, and putting a uniform distribution over response times, has no strong motivation in psychological theory. It is consistent with assumptions made by previous authors (e.g. Ratcliff & Tuerlinckx, 2002), and has some quali-

tatively desirable features, but there are many other forms for the prior that could reasonably be assumed. This issue highlights a basic challenge in contaminant identification and removal, which is that the best models will be those that have good justification for the contaminant data generating processes they assume. Accordingly, in our remaining two examples, we give examples of how contaminant models might be motivated by psychological theory, or come from previous modeling work.

Example 3: A Signal Detection Model of Recognition

Our third example deals with a standard Signal Detection Theory (SDT) model of recognition memory, embedded in a hierarchical model to account both individual and group behavior. There are two key points made by this example. The first is that the latent assignment approach to contaminant model applies in the same way for more *richly structured hierarchical models of cognition*. The second is that *psychological theory can make a contribution to defining contaminant processes*, by suggesting meaningful hypotheses about how people might fail to perform a task.

Background

Recognition memory tasks are among the most basic and widely used in the study of human memory. In these tasks, participants are presented a list of study words, and then presented with list of test words—containing both ‘old’ (target, signal) words from the study list and ‘new’ (distractor, lure, noise) words—and asked to identify the old words. Performance on the task is conveniently represented in terms of two counts: ‘hits’, corresponding to the number of recognized words in the study list, and ‘false alarms’, corresponding to the number of recognized words not in the study list.

A popular model of recognition memory uses equal-variance SDT (Green & Swets, 1966; MacMillan & Creelman, 2004), and accounts for the hit and false alarm counts using a combination of simple representation and decision-making assumptions. Representationally, it is assumed memory for each word can be summarized by a single ‘recognition strength’ value, and that these recognition strengths for studied and unstudied words are independent draws from Gaussian distributions with different means, but equal variances. The difference between the means is a measure of memory discriminability, and is usually denoted d' . The decision-making assumption of the model is that, if the recognition strength of a presented word exceeds some criterion value k , the participant will respond ‘yes’, but otherwise they will respond ‘no’. The difference between the criterion and ‘optimal’ unbiased threshold $d'/2$ provides a measure of decision-making bias, and is usually denoted c .

Most often, this SDT model is applied to aggregated count data across experimental groups, or to individual participants counts, to estimate the discriminability and bias parameters (e.g., Miller & Lewis, 1977; Snodgrass & Corwin, 1988; Yonelinas, Dobbins, Szymanski, Dhaliwal, & King, 1996). More recently, however, the model has been applied hierarchically, so that both individual and group parameters

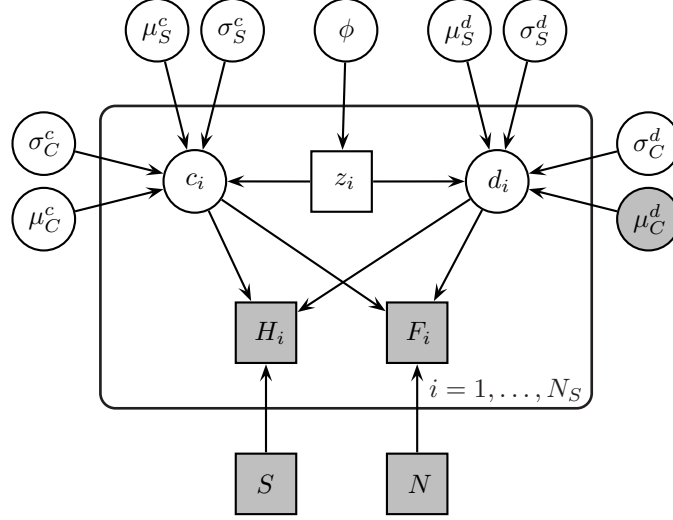


Figure 8. Graphical model for hierarchical signal detection analysis of recognition memory.

can be inferred (e.g., Dennis, Lee, & Kinnell, 2008; Rouder & Lu, 2005). Our example uses this extended hierarchical version, to demonstrate that latent assignment contaminant modeling applies in the same way to richly structured cognitive models.

Model

The standard equal-variance SDT model defines the number of hits for the i th subject, H_i as being binomially distributed from S target words with rate $h_i = \Phi(\frac{1}{2}d'_i - c_i)$, and the number of false alarms, F_i as binomially distributed from N noise words with rate $f_i = \Phi(-\frac{1}{2}d'_i - c_i)$. We assume, following Rouder et al. (2003), that within the discriminability and bias for each participant are Gaussian draws from an appropriate group distribution. To allow for contaminants, we define two of these groups, with means and standard deviations (μ_S^c, σ_S^c) and (μ_S^d, σ_S^d) , respectively, for the substantive group and (μ_C^c, σ_C^c) and (μ_C^d, σ_C^d) for the contaminant group.

We place assumptions about substantive and contaminant participants at the group level. Just as we considered “substantive” and “contaminant” to be individual designations in previous sections, we can likewise consider the collections of substantive and contaminant individuals, termed the substantive and contaminant groups. The psychological theory behind the SDT model parameters suggests one reasonable contaminant model would simply assume a discriminability value near zero. This corresponds to participants who are not engaged in the task, have not remembering the study words, and so are unable to separate their representation from the test words. Thus, for the contaminant group, we set $\mu_C^d = 0$. We define the \mathbf{Z} and ϕ parameters controlling the identification of contaminant individuals as in previous examples.

It remains to define appropriate prior distributions over the parameters for the

four group level distributions. In all cases, we use flat priors over model parameters. For the groups means, this is realized by a zero mean Gaussian whose standard deviation, ζ , we allow to go to ∞ . For the standard deviations, this is achieved by a gamma distribution whose two parameters are equal and themselves equal to ϵ and then allowing ϵ to go to zero. This results in the following model.

$$\begin{aligned}
H_i &| h_i, S \sim \text{Binomial}(S, h_i) \\
F_i &| f_i, N \sim \text{Binomial}(N, f_i) \\
c_i &| (\mu_S^c, \sigma_S^c), (\mu_C^c, \sigma_C^c), z_i \sim \begin{cases} \text{Gaussian}(\mu_C^c, \sigma_C^c), & \text{if } z_i = 0 \\ \text{Gaussian}(\mu_S^c, \sigma_S^c), & \text{if } z_i = 1 \end{cases} \\
d'_i &| (\mu_S^d, \sigma_S^d), \sigma_C^d, z_i \sim \begin{cases} \text{Gaussian}(0, \sigma_C^d), & \text{if } z_i = 0, \\ \text{Gaussian}(\mu_S^d, \sigma_S^d), & \text{if } z_i = 1 \end{cases} \\
z_i &| \phi \sim \text{Bernoulli}(\phi) \\
\mu_S^c, \mu_C^c, \mu_S^d &\sim \text{Gaussian}(0, \zeta) \\
\sigma_S^c, \sigma_C^c, \sigma_S^d, \sigma_C^d &\sim \text{Gamma}(\epsilon, \epsilon)
\end{aligned}$$

for the i th participant, with the constants $\zeta \rightarrow \infty$ and $\epsilon \rightarrow 0$. The graphical model is shown in Figure 8, and shows that the individual participant bias and discriminability parameters are conditionally independent given the group means and the substantive versus contaminant group assignment for each participant.

Analysis

We applied the model to unpublished data provided by Simon Dennis, in which 60 participants each completed 20 signal trials and 20 noise trials at test. The posterior distribution was sampled numerically using MCMC. For the simulations ζ was set to 0.001 and ϵ was set to 0.01. As with previous examples, four chains with random initial values were used, and the \hat{R} measure was used to check convergence. A burn-in of 100,000 samples was used, after which 5000 samples were drawn.

Figure 9 illustrates that two groups are present in the data, one with mean discriminability near 0, the other with mean discriminability near 3. This figure shows the posterior means of c_i and d'_i for each subject superimposed over a scatterplot of the joint posterior predictive distribution over new (c_i, d'_i) pairs for the latent assignment model. The posterior mode over the space of potential substantive versus contaminant assignments, $\hat{\mathbf{z}}$, is also shown. Means marked with triangles are assigned to the substantive model, whiel marked with diamonds are assigned to the contaminant.

The joint posterior distribution over potential substantive versus contaminant assignments \mathbf{Z} is summarized in Figure 10. This figure truncates posterior distribution over potential configurations by showing only those whose estimated posterior mass is greater than 10^{-3} . Thus, from a space of 2^{60} candidate assignments, after inference only nine assignments have posterior mass greater than 10^{-3} . One of these nine assignments, one assignment has mass greater than 0.93, the posterior mode $\hat{\mathbf{z}}$, which

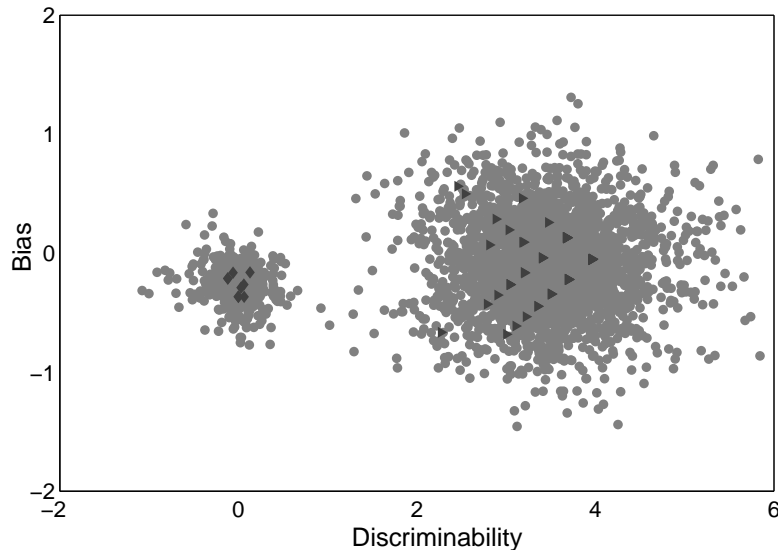


Figure 9. Posterior predictive distribution over d' , the discriminability, and c , the criterion value, for the hierarchical SDT models. Those points shown with triangles are assigned to the substantive group in the posterior mode over assignments and those with diamonds to the contaminant.

assigns participants 1, 7, 9, 20, 33, 36, 39, 45 and 57 to the contaminant group and all others to the substantive group. The other eight assignments are simply variations on the mode, changing $\hat{\mathbf{z}}$ by changing \hat{z}_i to $1 - \hat{z}_i$ for a single participant i . Each of these configurations is labeled in Figure 10 by the participant that is changed. For example, ‘-33’ moves participant 33 from the contaminant group to the substantive and ‘+24’ moves participant 24 from the substantive group to the contaminant.

As these results show, the latent assignment approach has lowered the space of potential assignments to handful with non-negligible mass and one highly probable assignment. This assignment indicates those participants who completed the recognition task with reasonable performance, and those whose performance is consistent with having zero discriminability, and so can meaningfully be interpreted as contaminants. Having identified the contaminant participants, in turn, allows for meaningful estimation of the substantive model parameters—discriminability and bias—for the group of participants shown in Figure 9 who did produce behavior consistent with motivated performance.

Example 4: Bandit Problem Decision-Making

Our fourth example involves a variety of heuristic models for understanding human behavior on a type of sequential decision-making problem known as a bandit problem (e.g., Sutton & Barto, 1998). For these problems, it is possible make different

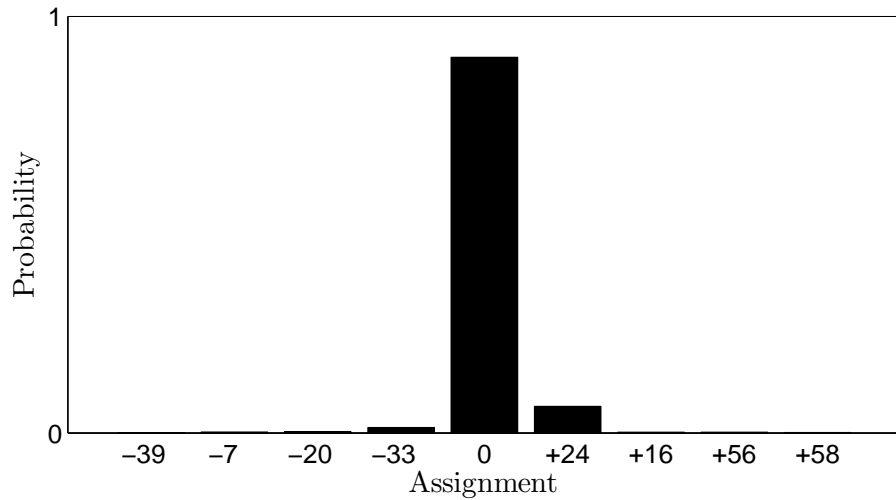


Figure 10. Probabilities of various group assignments. The standard, labeled '0', assigns participants 1, 7, 9, 20, 33, 36, 39, 45 and 57 to the contaminant group and all others to the substantive group. The other assignments are generated by adding '+' or subtracting '-' participants from the contaminant group, and placing them in the non-contaminant group.

reasonable assumptions about what constitutes the substantive model of interest, and what models are used to detect contaminant behavior. The key contributions of the example is to demonstrate that the *details of the assumptions about contamination affect the inferences made about the model being studied*, and so finding a good set of contaminant models can often be a necessary and significant research undertaking in its own right.

Background

In bandit problems, a decision-maker chooses repeatedly between a set of alternatives. They get feedback after every decision, either recording a reward or a failure. They also know that each alternative has some fixed, but unknown, probability of providing a reward each time it is chosen. The goal of the decision-maker is to obtain the maximum number of rewards over all the trials they complete. In one popular version of the bandit problem, known as the finite horizon problem, the number of trials is fixed, known, and usually small.

Studies of human decision-making on bandit problems have been approached from a variety of theoretical perspectives, including operant conditioning (e.g., Brand, Wood, & Sakoda, 1956; Brand, Sakoda, & Woods, 1957), rationality in economic decision-making (e.g., Anderson, 2001; Banks, Olson, & Porter, 1997; Horowitz, 1973; Meyer & Shi, 1995), and, most recently, computational cognitive science (e.g. Steyvers, Lee, & Wagenmakers, 2009) and cognitive neuroscience (e.g., Cohen, McClure, & Yu, 2007; Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006).

One interesting issue in studying human performance on bandit problems involves the potential use of different decision-making strategies. There are many psychologically plausible heuristic approaches coming from the game theory and reinforcement learning literatures (e.g. Sutton & Barto, 1998), as well as heuristics developed in the cognitive sciences (e.g. Zhang, Lee, & Munro, 2009), and there is some empirical evidence that different people use different heuristics in the same experiment (Steyvers et al., 2009). Some of these heuristics are quite sophisticated, and represent what might be viewed as intelligent or effective approaches. Others are very simple, and clearly sub-optimal.

This contrast raises the issue of exactly what constitutes “contaminant” behavior in a bandit problem experiment. If the focus is on understanding the relatively sophisticated models by, for example, inferring model parameters from behavioral data, then the simple heuristic approaches can be viewed as contaminating. In this section, we present some analyses of bandit problem decision-making showing how different assumptions about what qualifies as contaminant behavior influences what is learned about the substantive model.

Model and Data

We choose a standard heuristic usually known as “Win-Stay Lose-Shift” (WSLS) as our substantive model (Robbins, 1952). It assumes that if, after choosing an alternative, the decision-maker is rewarded, they will choose the same alternative on the next trial with some (high) probability γ . Alternatively, if the decision-maker is not rewarded, WSLS assumes they will only choose the same alternative on the next trial with some (small) probability $1 - \gamma$.

While extremely simple, the WSLS often provides a reasonable account of people’s decision-making. For example, Steyvers et al. (2009) collected data from 451 participants on a series of bandit problems, and presented a series of model comparisons showing that the majority of these participants decisions consistent with WSLS. We use an abbreviated version of the same data set—using a subset of participants chosen to make clear the contaminant modeling principles this example aims to explain—including 47 participants. As with the full data set, all participants completed a set of 20 bandit problems, each involving four alternatives and 15 trials.

This means that the substantive WSLS model assumes the choice of the k th participants on the t th trial of the g th problem, $d_{k,g,t}$, given the presence or absence of reward on the preceeding trial, $r_{k,g,t-1}$ is

$$\Pr(d_{k,g,t} = i \mid M_{\text{WSLS}}) = \begin{cases} \gamma & \text{if } d_{k,g,t-1} = i \text{ and } r_{k,g,t-1} = 1 \\ 1 - \gamma & \text{if } d_{k,g,t-1} = i \text{ and } r_{k,g,t-1} = 0 \\ \frac{1}{3}(1 - \gamma) & \text{if } d_{k,g,t-1} \neq i \text{ and } r_{k,g,t-1} = 1 \\ \frac{1}{3}\gamma & \text{if } d_{k,g,t-1} \neq i \text{ and } r_{k,g,t-1} = 0. \end{cases}$$

For the bandit problem, it is easy to think of at least two plausible strategies a non-motivated participant might use to complete the task. One, we call the ‘random’

strategy, in which they simply chose an alternative at random on every trial. Under this model,

$$\Pr(d_{k,g,t} = i \mid M_{\text{RAND}}) = \frac{1}{4}$$

for all alternatives, on every trial of every game. A second non-motivated strategy we call ‘same’, and involves the participant choosing the same alternative on almost every trial, regardless of the observed pattern of reward. If the i^* th alternative is the favored one, a simple formal implementation of this strategy is given by

$$\Pr(d_{k,g,t} = i \mid M_{\text{SAME}}) = \begin{cases} 0.95 & \text{if } i = i^* \\ 0.05 & \text{otherwise.} \end{cases}$$

Results

We applied the three models—the substantive WSLs, and the contaminant random and same heuristics—to the Steyvers et al. (2009) data in four separate analyses. In the first, we simply applied the WSLs model. In the second analysis, we applied WSLs, but also introduced the random model as a contaminant model, using the latent assignment approach. In the third analysis we applied WSLs with the same model as the contaminant model. In the fourth analysis, we used both the random and same models as contaminants, allowing the behavior of each participant to be explained by any one of these three accounts.

The graphical model for just the fourth analysis—which incorporates the new development of allowing two possible contaminant models—is shown in Figure 11. The probability the i th alternative is chosen on the t th trial of the g th game by the k th participant is given by $\theta_{k,g,t}^i$. Which of the three models is used to determine this probability is decided by the three-valued latent assignment variable z_k . This variable is controlled by two latent base-rates, ϕ and η . The base-rate ϕ controls the probability over all participants that the substantive WSLs model is used. The latent base-rate η controls the probability, for those participants not using WSLs, that they follow the random or same contaminant model.

This means that $z_k \sim \text{Categorical}(\phi, (1 - \phi)\eta, (1 - \phi)(1 - \eta))$, so that

$$\theta_{k,g,t}^i = \begin{cases} \Pr(d_{k,g,t} = i \mid M_{\text{WSLS}}) & \text{if } z_k = 1 \\ \Pr(d_{k,g,t} = i \mid M_{\text{RAND}}) & \text{if } z_k = 2 \\ \Pr(d_{k,g,t} = i \mid M_{\text{SAME}}) & \text{if } z_k = 3. \end{cases}$$

When WSLs is used, the model uses the rate γ and whether or not a reward was obtained for the i th alternative on the previous trial, $r_{k,g,t-1}^i$. These dependencies are also shown in the graphical model in Figure 11. We assumed uniform prior distributions over the latent assignment rates ϕ and η , and also over the WSLs rate γ .

The left panel of Figure 12 shows how the participants were assigned to the three models, according to the mode of the posterior for each z_k . Each point corresponds to a participant, and the type of marker indicates whether they were classified

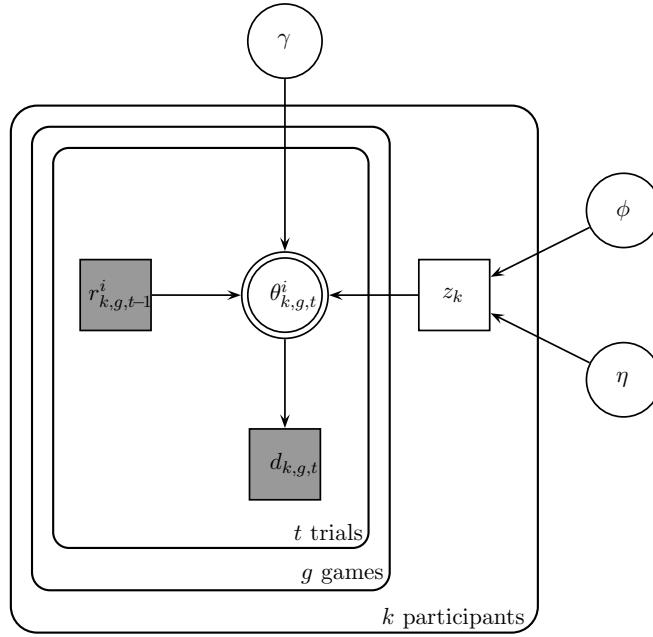


Figure 11. Graphical model for the analysis of bandit problem decision-making using WSLs as a substantive model, and the random and same strategies as contaminant models.

as following the WSLs, random or same model. The axes in which the points are displayed correspond to two summary measures of their decision-making, chosen because they capture much of the variance involved in partitioning the participants among the models. The x -axis shows the proportion of trials following no reward that a different alternative was chosen on the next trial. The y -axis shows the proportion of trials following a reward the same alternative was chosen on the next trial.

WSLS performance corresponds to high values on both measures, and so these participants are in the top-right corner. Random model performance corresponds to the point (0.75, 0.25), since there are four alternatives. Same model performance correspond to the top-left corner of the graph. The left panel of Figure 12 shows a clear partitioning of participants into each of these regions, and that they are appropriately assigned by the model. In other words, there are clear individual differences between participants in the decision strategy these use to solve bandit problems, and they appear to be well described by the WSLs, random and same models for these participants.

The inferences about the γ parameter of WSLs are shown, for all four analyses, in the right panel of Figure 12. The key point is that the inferred rate of winning and staying or losing and shifting changes significantly depending on the assumptions made about contaminant behavior in the participant pool. When no contamination is assumed, γ is around 0.75. When both the same and random forms of contamination

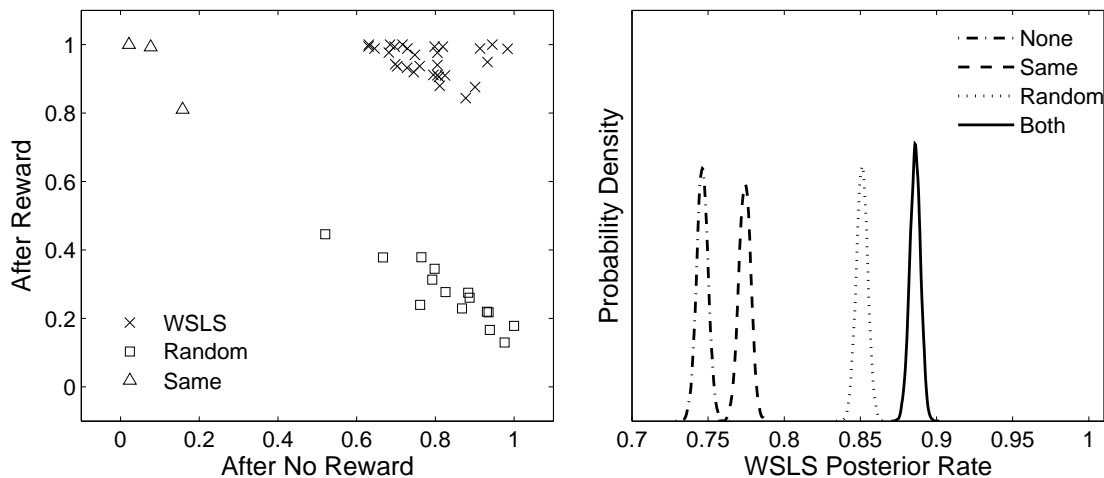


Figure 12. Analysis of bandit problem behavior. The left panel shows the 47 participants, and their assignment to the WSLs, random and same models. The right panel shows the posterior distribution of the WSLs rate γ for four analyses, including no contaminant modeling, the random contaminant model, the same contaminant model, or both contaminant models.

are included in the analysis, the inferred γ increases to almost 0.9. Using just one or other of the contaminant models gives different intermediate values. These results make clear that what is learned by applying a substantive cognitive model to behavioral data can depend critically on the nature of possible contamination processes included in the analysis.

Conclusion

In our four worked examples, we have tried to argue for a latent assignment approach to dealing with contaminants in psychological data. The basic assumption of the approach is that a good cognitive model should aim to account for all of the observed behavior, but recognize that not all experimental data will be generated by the cognitive processes that are of theoretical interest. The latent assignment approach allows one or more models of contaminant behavior to be mixed with the substantive model of interest, within a coherent Bayesian framework for inference.

Our examples show that there is no single principled method that can define appropriate contaminant models automatically. In fact, we believe that modeling contaminant data involves the same creative exercise as modeling substantive psychological processes. Just as we do not know how to automate the building of cognitive models for memory, learning, decision-making, and other cognitive processes, we do not know how to automate the non-compliant cognitive behaviors that generate contaminants in experimental tasks. Rather, building contaminant models should be

treated as a core modeling problem, and pursued by attempting to find a statistical characterization or psychological model that provides a good account of the data produced when people are not behaving as intended.

What our examples also show, however, is that the latent assignment approach to mixture modeling provides a general and coherent framework for introducing proposed contaminant models. Given a useful contaminant model, the framework allows contaminant data to be identified, and removes their influence on parameter estimation and model selection. It achieves this in a way that naturally incorporates the base-rate of contaminants in a data set, and is applicable to contaminants existing at the level of trials, participants, groups, or at any other level. In addition, the latent assignment method applies in the same way to any formal probabilistic model of cognition, including richly structured hierarchical models. We think the potential problems posed by contaminant data, coupled with all of these attractive features of the latent assignment approach, should encourage its widespread use in cognitive modeling.

Acknowledgements

This work was funded by award FA9550-07-1-0082 from the Air Force Office of Scientific Research. We thank two anonymous reviewers for their comments on an earlier version of this paper.

References

- Anderson, C. M. (2001). *Behavioral models of strategies in multi-armed bandit problems*. Unpublished doctoral dissertation, California Institute of Technology.
- Banks, J., Olson, M., & Porter, D. (1997). An experimental analysis of the bandit problem. *Economic Theory*, 10, 55–77.
- Barnett, V. (1978). The study of outliers: Purpose and model. *Applied Statistics*, 27, 242–250.
- Box, G. E. P. (1979). Some problems of statistics and everyday life. *Journal of the American Statistical Association*, 74, 1–4.
- Brand, H., Sakoda, J. M., & Woods, P. J. (1957). Effects of a random versus pattern reinforcement instructional set in a contingent partial reinforcement situation. *Psychological Reports*, 3, 473–479.
- Brand, H., Wood, P. J., & Sakoda, J. M. (1956). Anticipation of reward as a function of partial reinforcement. *Journal of Experimental Psychology*, 52(1), 18–22.
- Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic cognition approach to decision making. *Psychological Review*, 100, 432–459.

- Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? exploration versus exploitation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362, 933–942.
- Daw, N. D., O’Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, 441, 876–879.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm (with discussion). *Journal of the Royal Statistical Society series B*, 39, 1–38.
- Dennis, S., Lee, M., & Kinnell, A. (2008). Bayesian analysis of recognition memory: The case of the list-length effect. *Journal of Memory and Language*, 59, 361–376.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis (2nd ed.)*. Boca Raton (FL): Chapman & Hall/CRC.
- Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*. London: Chapman and Hall.
- Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, 78(1), 27–43.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Horowitz, A. D. (1973). *Experimental study of the two-armed bandit problem*. Unpublished doctoral dissertation, The University of North Carolina, Chapel Hill, NC.
- Jaynes, E. T. (2003). *Probability theory: The logic of science*. New York: Cambridge University Press.
- Jordan, M. I. (2004). Graphical models. *Statistical Science*, 19, 140–155.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2004). *Applied linear statistical models* (5 ed.). McGraw-Hill/Irwin.
- Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review*, 12(1), 1–15.
- Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, 40, 77–105.
- Little, D. R., & Lewandowsky, S. (2009). Beyond non-utilization: Irrelevant cues can gate learning in probabilistic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 530–550.
- Little, D. R., Lewandowsky, S., & Heit, E. (2006). Ad hoc category restructuring. *Memory & Cognition*, 34, 1398–1413.

- Lunn, D. J., Spiegelhalter, D., Thomas, A., & Best, N. (in press). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*.
- Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS – a Bayesian modelling framework: Concepts, structure and extensibility. *Statistics and Computing*, 10, 325–337.
- MacMillan, N., & Creelman, C. D. (2004). *Detection theory: A user's guide (2nd ed.)*. Hillsdale, NJ: Erlbaum.
- Meyer, R. J., & Shi, Y. (1995). Sequential choice under ambiguity: Intuitive solutions to the armed-bandit problem. *Management Science*, 41(5), 817–834.
- Miller, E., & Lewis, P. (1977). Recognition memory in elderly with depression and dementia: A signal detection analysis. *Journal of Abnormal Psychology*, 86(1).
- Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, 104, 266–300.
- Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press.
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 20, 873–922.
- Ratcliff, R., & Rouder, J. F. (2000). Modeling response times for two-choice decisions. *Psychological Science*, 9, 347–356.
- Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, 9(3), 438–481.
- Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 55, 527–535.
- Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, 12(4), 573–604.
- Rouder, J. N., Sun, D., Speckman, P. L., Lu, J., & Zhou, D. (2003). A hierarchical Bayesian framework for response time distributions. *Psychometrika*, 68, 589–606.
- Rudas, T., Clogg, C. C., & Lindsay, B. G. (1994). A new index of fit based on mixture methods for the analysis of contingency tables. *Journal of the Royal Statistical Society: Series B*, 56, 623–639.
- Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology*, 44, 408–463.
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117(1), 34–50.

- Steyvers, M., Lee, M. D., & Wagenmakers, E. (2009). A Bayesian analysis of human decision-making on bandit problems. *Journal of Mathematical Psychology*, 53, 168–179.
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge (MA): The MIT Press.
- Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82, 528–540.
- Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2008). A Bayesian approach to diffusion process models of decision-making. In V. Sloutsky, B. Love, & K. McRae (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1429–1434). Austin, TX: Cognitive Science Society.
- Vickers, D. (1979). *Decision processes in visual perception*. New York: Academic.
- Yonelinas, A. P., Dobbins, I., Szymanski, M. D., Dhaliwal, H. S., & King, L. (1996). Signal-detection, threshold, and dual-process models of recognition memory: ROCs and conscious recollection. *Consciousness and Cognition*, 5(4), 418–441.
- Zacks, J. A. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, 28(6), 979–1008.
- Zeithamova, D., & Maddox, W. T. (2006). Dual task interference in perceptual category learning. *Memory & Cognition*, 34, 387–398.
- Zhang, S., Lee, M. D., & Munro, M. N. (2009). Human and optimal exploration and exploitation in bandit problem. In A. Howes, D. Peebles, & R. Cooper (Eds.), *Proceedings of the Ninth International Conference on Cognitive Modeling — ICCM2009*. Manchester, UK.