Three Bayesian Analyses of Memory Deficits in Patients with Dissociative Identity Disorder

Michael D. Lee
University of California, Irvine

Tom Lodewyckx
University of Amsterdam

E.–J. Wagenmakers
University of Amsterdam

Abstract

We demonstrate the flexibility and usefulness of Bayesian inference by modeling data on memory processes in patients with Dissociative Identity Disorder (DID). DID patients often report impaired memory for events that have been experienced by other personalities. For instance, the personality "plumber Joe" may have limited or no knowledge of the events experienced by the personality "gambler Mike". In order to assess whether or not DID patients simulate memory impairment, Huntjens et al. (2006) carried out a potentially diagnostic experiment featuring healthy participants who were either never shown the study materials (i.e., Amnesiacs) or who were told to simulate impairment (i.e., Malingerers). We re-analyze their data in three different ways using Bayesian statistical methods. Each analysis supports the hypothesis that the data do not allow one to conclusively answer the original question. Our conclusion from the Bayesian analyses is that the performance of DID-patients is about as similar to that of Amnesiacs as it is to that of Malingerers, and the DID patients may well constitute a separate group.

Introduction

One of the hallmarks of Rich Shiffrin as a scientist, especially as the audience member asking the first question after a research presentation, is his instinctive skepticism towards definitive claims. Asserting that some psychological conclusion follows from the application of a cognitive model to data is likely to be met with counter-arguments that force the careful qualification of the claim. Rich has at the front of his mind that human cognition is extremely complicated, and the models we use are at best crude approximations to that complexity. Models are always wrong, but some can be useful, if used and interpreted in a careful and critical way.

In this chapter, we consider a research question involving memory processes in patients with Dissociative Identity Disorder (DID, also known as Multiple Personality Disorder), based on a study reported by Huntjens et al. (2006). DID patients often report *inter-identity amnesia*, that is, impaired memory for events experienced by personalities that are not currently present. The research question is how the memory performance of DID patients in a simple recognition task relates to the performance of people from various control and comparison experimental conditions.

Taking our lead from Rich's wariness and skepticism, we approach the same research questions from three different complementary perspectives. We conduct three separate Bayesian analyses of the same data, considering the memory performance of the DID patients from both parameter estimation and model selection perspectives. The models we use improve on the original frequentist analysis of the data in a number of ways, including by allowing for individual differences (e.g., Shiffrin, Lee, Kim, & Wagenmakers, 2008). While each model remains a drastic simplification of the richness of human memory in a complicated clinical setting, each analyses does support the same substantive conclusion.

Huntjens et al. (2006) Study

To test whether DID-patients were really affected by inter-identity amnesia, or whether they were simulating their affliction, Huntjens et al. (2006) conducted a clever experiment that assessed the performance of four groups of participants on a multiple-choice recognition test. In the learning stage of the task, participants were

Address correspondence to Michael D. Lee, Department of Cognitive Sciences, School of Social Sciences, University of California, Irvine. Email: mdlee@uci.edu.

presented with a brief story and several drawn figures. Later, in the testing stage of the task, participants were given a multiple choice recognition test consisting of 15 items: 10 of these items (each with 3 choice options) tested knowledge of story details, and the remaining 5 items (each with 5 choice options) required the identification of the drawn figures.

The first experimental group consisted of 19 *DID Patients*, who were asked to switch identities between studying the materials and being tested on the materials. The second group, known as *Malingerers*, were 25 healthy adults instructed to simulate inter-identity amnesia, based on detailed information they were given about DID. In other words, they were instructed to simulate a switch of identities between the study and test phases. The third group, known as *Amnesiacs*, were 25 healthy adults who were never shown the study list and therefore truly had no knowledge of the relevant material during testing. The fourth *Control* group were 25 healthy adults instructed to perform the entire task as accurately as possible.

If DID patients are truly affected by inter-identity amnesia, their performance should resemble that of the Amnesiacs (i.e., participants who had not seen the study materials); if DID patients are simulating their affliction, their performance should resemble that of the Malingerers (i.e., participants told to simulate inter-identity amnesia). In terms of the number of questions answered correctly, Huntjens et al. (2006) found means and standard deviations of $M = 1.88$, $SD = 1.59$ for the Malingerers; $M = 3.11$, $SD = 1.59$ for the DID Patients; $M = 4.56$, $SD = 1.83$ for the Amnesiacs; and $M = 13.28$, $SD = 1.46$ for the Controls.

These data indicate that the Controls performed best, and that the DID Patients performed better than the Malingerers, but worse than the Amnesiacs. Huntjens et al. (2006) observed that the DID patients are about halfway between the Malingerers and the Amnesiacs, and suggested they may constitute a separate group. They also reported, however, that the performance of DID Patients was significantly worse than that of Amnesiacs ($p = .03$), whereas "Importantly, patients did not differ significantly from simulators ($p = 0.09$)." (Huntjens et al., 2006, p. 5). This line of reasoning suggests that the DID patients are more similar to the Malingerers than to the Amnesiacs. However, the difference between "significant" and "not significant" is itself not necessarily statistically significant (Gelman & Stern, 2006; see also Nieuwenhuis, Forstmann, & Wagenmakers, 2011); hence, without additional analyses one cannot conclude that performance of DID patients resembles that of Malingerers

more than that of Amnesiacs.

Thus, while the clinical question about memory in DID Patients is important, and the experiment was well conceived and executed, the reported analysis does not convincingly answer the research question. It is not clear whether the data provide evidence that DID Patients behave in a way consistent with inter-identity amnesia, or consistent with simulation, or in a way consistent with being different from both of these possibilities.

Huntjens et al. (2006) Raw Data

The relevant experimental data from the Huntjens et al. (2006) study consist of the number of correct recognition responses for each individual participant in each group. These raw data were not made available to us. However, it is possible to make inferences about the possible values of the raw data from published summaries. Specifically, Klugkist (2008, Table 4.1) details the number of correct responses for the individual DID Patients and Amnesiacs. Furthermore, Klugkist (2008, Table 4.4) gives the means and standard deviations of correct responses for each of the groups, as detailed above. Finally, Rossell, Baladandayuthapani, and Johnson (2008, Figure 6.1) presents a standard box-plot graphical summary of all four groups, indicating the upper and lower bounds on the number of correct responses in each group.

We used this available information in order to infer the raw values for the Malingering and Control groups. We did this by considering all possible data sets that adhered to the upper and lower bounds with matching means and standard deviations. For all data sets that met these constraints, we visually examined the box-plots, retaining only those that matched (Rossell et al., 2008, Figure 6.1). The results of this procedure are shown in Table 1, which lists all of the possibilities for the raw data. For the DID Patients and the Amnesiacs, the numbers of correct recognitions are listed. For both Malingerers and Controls, there turned out to be five possibilities consistent with the constraints, and these are listed. Combining the five possibilities for the Malingerers with the five possibilities for the Controls, this means there are a total of 25 possible raw data sets.

Three Bayesian Analyses

Using the inferred raw data from Huntjens et al. (2006), we undertook three analyses that examined, from different perspectives, how the DID Patients relate to

Table 1: Inferred raw data from the experiment reported by Huntjens et al. (2006). For the DID Patients and Amnesiacs, the known numbers of correct recognitions for each individual are listed. For both Malingerers and Controls, the five possible data sets are listed as PD-1 through PD-5.

DID Patients																									
Data	0	1	2	2	2	2	3	3	3	3	3	3	3	4	4	4	4	6	7						
Amnesiacs																									
Data	1	1	2	3	3	4	4	4	4	4	4	4	5	5	5	5	5	5	5	6	6	6	6	8	9
Malingerers																									
PD-1	0	0	0	0	0	0	1	1	1	1	1	1	2	2	2	2	2	3	3	4	4	4	4	4	5
PD-2	0	0	0	0	0	0	1	1	1	1	1	1	2	2	2	2	3	3	3	3	3	4	4	5	5
PD-3	0	0	0	0	0	0	1	1	1	1	1	2	2	2	2	2	2	2	3	3	4	4	4	5	5
PD-4	0	0	0	0	0	0	1	1	1	1	1	2	2	2	2	2	2	3	3	3	3	3	5	5	5
PD-5	0	0	0	0	0	1	1	1	1	1	1	1	2	2	2	2	2	2	3	3	3	4	5	5	5
Controls																									
PD-1	9	10	12	12	12	12	12	13	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	15	15
PD-2	9	11	11	12	12	12	12	13	13	14	14	14	14	14	14	14	14	14	14	14	14	14	14	15	15
PD-3	9	11	11	12	12	12	13	13	13	13	13	14	14	14	14	14	14	14	14	14	14	14	15	15	15
PD-4	9	11	12	12	12	12	12	12	13	13	14	14	14	14	14	14	14	14	14	14	14	14	15	15	15
PD-5	9	11	12	12	12	12	12	13	13	13	13	13	14	14	14	14	14	14	14	14	14	15	15	15	15

the Amnesiacs and Malingerers. All three analyses use a hierarchical modeling approach that allows for individual differences, and rely on Bayesian inference with its guarantees of completeness, coherence, and consistency (e.g., Shiffrin et al., 2008). All analyses were carried out in WinBUGS (Lunn, Jackson, Best, Thomas, & Spiegelhalter, 2012) and JAGS (Plummer, 2003), programs that facilitate Bayesian inference through the use of Markov chain Monte Carlo sampling (Lee & Wagenmakers, 2013).

As detailed below, the first analysis provides the posterior distribution of the key difference in group means. The second analysis provides the posterior distribution of the rate with which DID patients are assigned to either the Amnesiacs or the Malingerers. The third analysis provides the Bayes factors for four possible clusterings of the groups.

Analysis 1: Posterior Difference in Group Mean Rates. The first analysis infers the posterior distribution of the difference between two quantities: the difference in mean success rate between DID Patients versus Malingerers and the difference in mean success rate between DID Patients versus Amnesiacs. If this distribution is

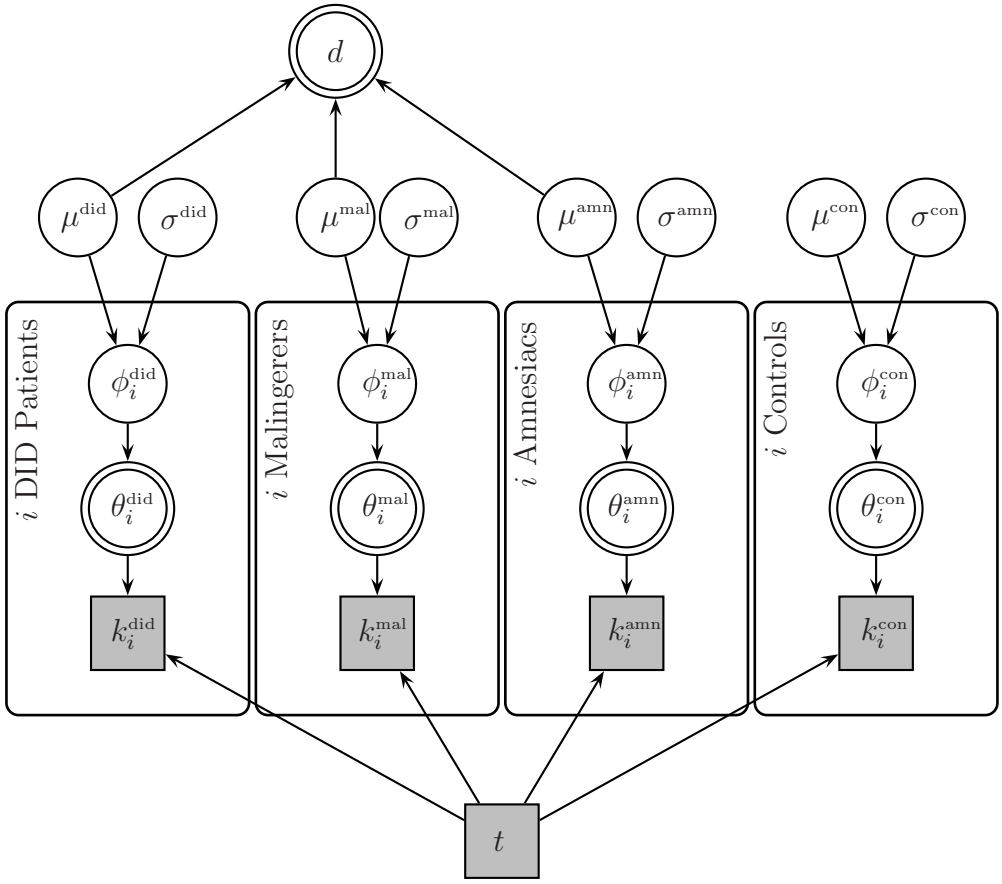

Figure 1. Graphical model for inferring the mean accuracy for each group, and the key difference d between the groups means for the Malingering, DID Patient and Amnesiac groups, such that $d = (\mu^{\text{amn}} - \mu^{\text{did}}) - (\mu^{\text{did}} - \mu^{\text{mal}})$.

symmetric and centered around zero, this indicates that DID Patients differ just as much from Malingerers as they do from Amnesiacs.

Figure 1 shows the model we developed to make this inference, using graphical model notation, which is a formalism for implementing probabilistic models of cognitive processes (Lee, 2011; Lee & Wagenmakers, 2013; Shiffrin et al., 2008). In graphical models, nodes represent variables and data, and the graph structure is used to indicate dependencies. Continuous variables are represented with circular nodes and discrete variables are represented with square nodes. Observed variables, which are usually data or properties of an experimental design, are shaded and unobserved variables, which are usually model parameters, are not shaded. Plates are square

boundaries that enclose subsets of the graph that have independent replications in the model.

In Figure 1, the discrete observed variable $t = 15$ represents the total number of items in the recognition test. The discrete observed variables k_i^g correspond to the data, representing the number of correct answers for the ith participant in the gth group. We assume that there is an underlying probability θ_i^g for this participant answering any question correctly, so that $k_i^g \sim \text{Binomial}\big(\theta_i^g, t\big)$.

To allow for individual differences in the accuracy rates, the model assumes the accuracy of each participant comes from a group level distribution. This approach means that participants in the same experimental group are expected to have similar accuracies, consistent with the possibility of meaningful differences between the groups, but allows for some individual variation, consistent with the assumption of individual differences. Statistically, this is accomplished using a probit transformation, which provides a link between Gaussian group distributions (that covers the entire real line) and individual probabilities (that cover the unit interval). Specifically, a sample ϕ_i^g is drawn from the group-level Gaussian with mean μ_i^g and standard deviation σ_i^g, and the accuracy is then found by the transformation $\theta_i^g = \Phi\big(\phi_i^g\big)$, where $\Phi\big(\cdot\big)$ is the cumulative standard Gaussian distribution function. This means the accuracy θ_i^g is a reparametrization of another variable, and so is a deterministic node in the graphical model, indicated by double borders. The model uses the priors $\mu_i^g \sim \text{Gaussian}\big(0, 1\big)$, which corresponds to a uniform prior on mean group accuracy, and a vague prior $\sigma_i^g \sim \text{Uniform}\big(0, 5\big)$ on the standard deviation.

Finally, the graphical model in Figure 1 calculates the deterministic variable $d = (\mu^{\text{amn}} - \mu^{\text{did}}) - (\mu^{\text{did}} - \mu^{\text{mal}})$. This variable indicates the extent to which the difference in performance, at a group level, between the Amnesiacs and DID Patients is bigger than the difference between the Malingerers and DID Patients.

Figure 2 shows the group means after transformation from the probit scale to the probability scale. All 25 possible data sets are analyzed, with one randomly chosen data set highlighted and labeled. All of the possible data sets lead to the same conclusion, which is that there is little overlap between the posterior distributions of group means.

Figure 3 shows the posterior distribution of d. Since the calculation of this difference does not depend on the Controls, there are only five possible data sets that need to be considered. A randomly chosen one is again highlighted and labeled. The

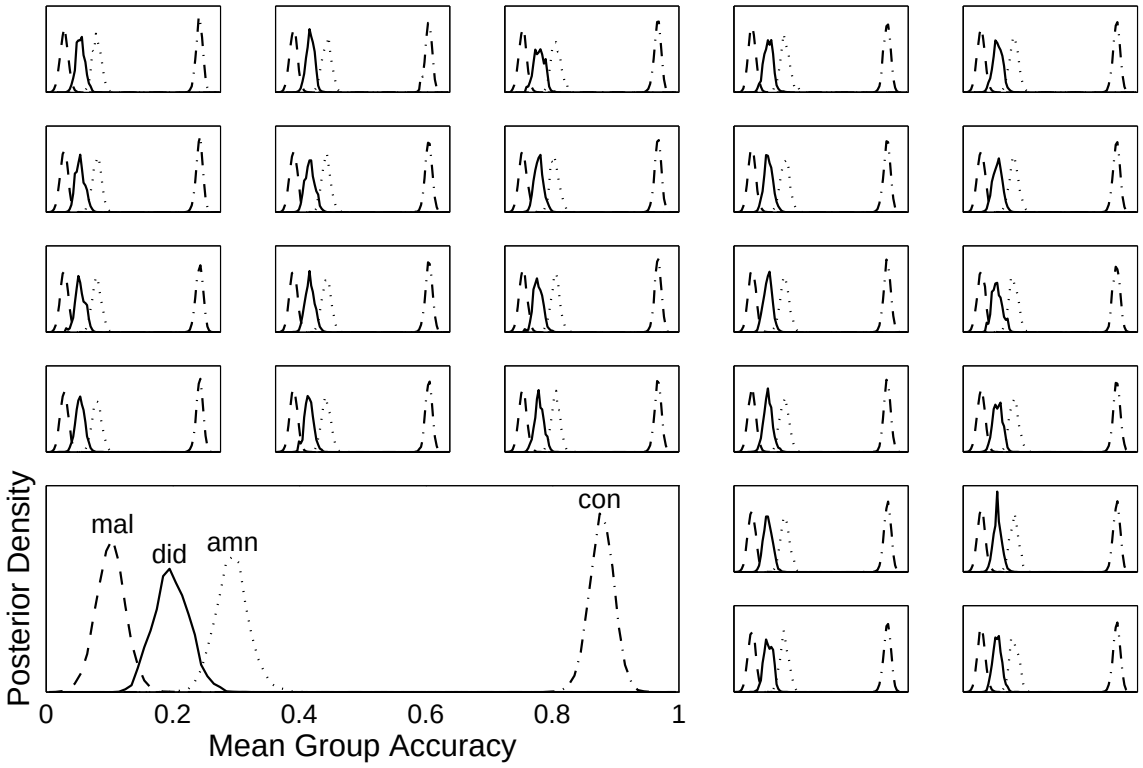

Figure 2. Posterior distribution of group means for each condition, for each of the 25 possible data sets. One data set, chosen at random, is shown enlarged and labeled.

distribution is symmetric and its mode is close to zero, indicating that the group mean rate for DID Patients is almost exactly in between that of Malingerers and Amnesiacs.

Analysis 2: Latent Assignment of DID Patients. The second analysis uses a different approach to inferring whether the DID Patients are closer to Malingerers or Amnesiacs. We assume that every DID Patient belongs to exactly one of these two groups, and then use a latent mixture model to infer their group membership.

The graphical model for this analysis is shown in Figure 4. The basic assumptions about individual differences and how responses are generated remain the same. The key difference is the introduction of a binary latent variable z_i for the ith DID Patient, which represents whether that patient is classified as a Malingerer or Amnesiac, by sampling their accuracy from one or other of those group distributions.

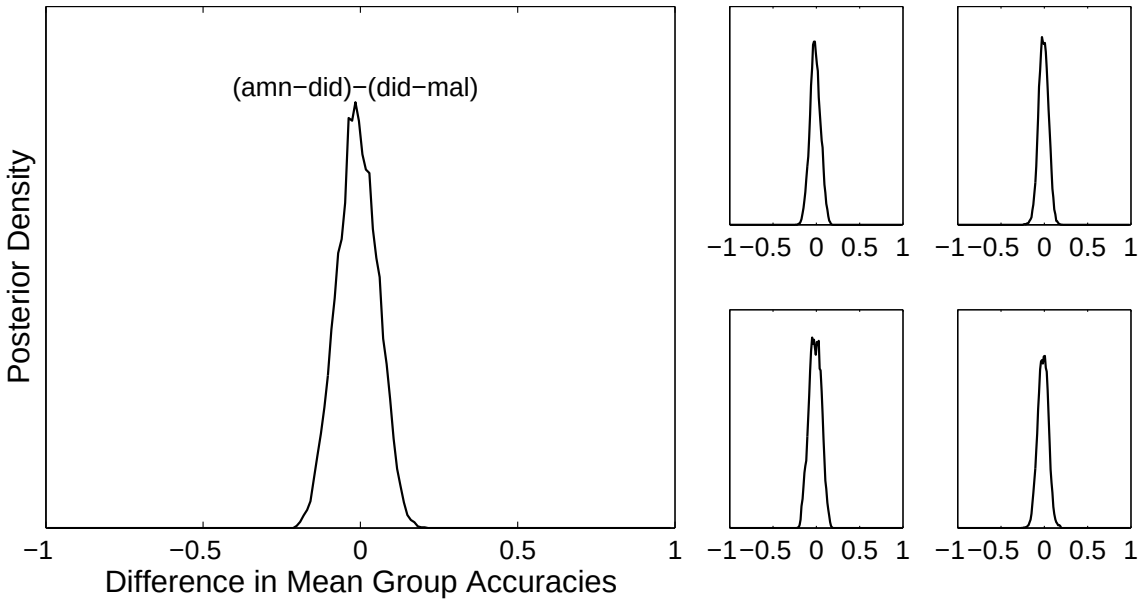

Figure 3. Posterior distribution of the key difference between the means. One data set, chosen at random, is shown enlarged and labeled.

Formally,

$$\phi_i^{\text{did}} \sim \begin{cases} \text{Gaussian}\big(\mu^{\text{mal}}, \sigma^{\text{mal}}\big) & \text{if } z_i = 0, \\ \text{Gaussian}\big(\mu^{\text{amn}}, \sigma^{\text{amn}}\big) & \text{if } z_i = 1. \end{cases}$$

A prior of $z_i \sim \text{Bernoulli}\big(\psi\big)$, corresponding to a base-rate $\psi \sim \text{Uniform}\big(0, 1\big)$ is placed on these classification variables.

Figure 5 shows the posterior expectation of the z_i variables for each DID Patient by circles, and the posterior expectation of the base-rate ψ by a broken line. These expectation correspond to the probabilities that individual participants, and the group as a whole, belong to the Amnesiacs versus Malingerers. Again, there are only five possible data sets, because this analysis does not involve the Controls. Inferences for all of the possibilities are shown, and the results for one randomly chosen data set are highlighted and labeled. While there are some differences between the possible data sets, the same basic conclusions hold in each case. For many DID patients, it is highly uncertain as to which group they belong, and the group base-rate is approximately halfway between the Amnesiac and Malingering groups. In addition, for those patients classified with some certainty into one of the groups, there are approximately as many classified as Amnesiacs as Malingerers. Overall, there is no evidence for believing the

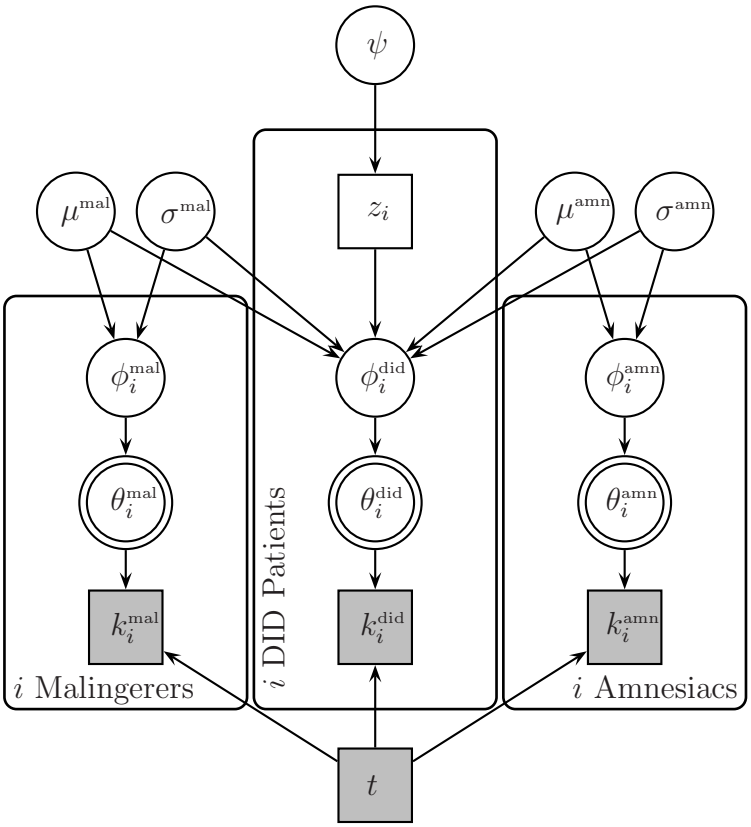

Figure 4. Graphical model for inferring the latent assignment of DID Patients to the Amnesiac or Malingering group.

DID Patients are more like either of the alternative groups.

Analysis 3: Bayes Factors Between Four Possible Clusterings. The third analysis seeks to distinguish four possible structures that correspond to different assumptions about the relationship between the DID Patient group and the Malingering and Amnesiac groups. The first structure is that all of these groups have the same mean, so that $\mathcal{S}_1 : \mu^{\text{mal}} = \mu^{\text{did}} = \mu^{\text{amn}}$. The second is that Malingerers and DID Patients are the same, but the Amnesiacs have a different greater mean, so that $\mathcal{S}_2 : \mu^{\text{mal}} = \mu^{\text{did}} < \mu^{\text{amn}}$. The third structure is that the DID Patients and Amnesiacs are the same, but the Malingerers have a different lower mean, so that $\mathcal{S}_3 : \mu^{\text{mal}} < \mu^{\text{did}} = \mu^{\text{amn}}$. The final structure assumes that all of the groups have different ordered means, so that $\mathcal{S}_4 : \mu^{\text{mal}} < \mu^{\text{did}} < \mu^{\text{amn}}$. These four possibilities span a range from the groups being completely the same to completely different, including as intermediate possibilities

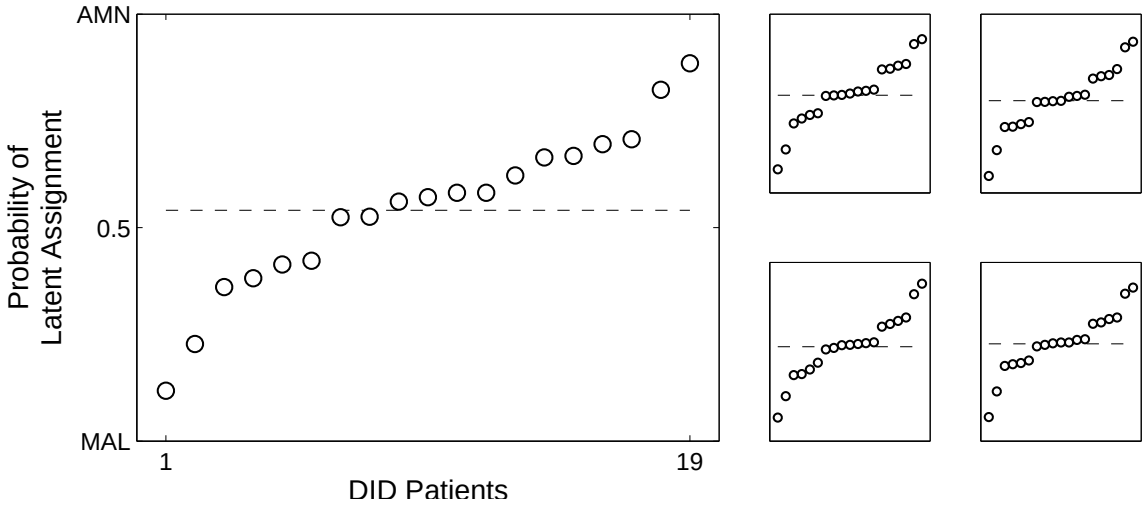

Figure 5. Expectation of latent assignments to the Malingering versus Amnesiac group for each DID Patient (circles), and the base-rate average of assignment (broken line). One data set, chosen at random, is shown enlarged and labeled.

that the DID Patient group is the same as either the Malingerers or Amnesiacs.

Figure 6 shows a graphical modeling framework that can accommodate all of the structures, and allows the calculation of the Bayes factors for their comparison (Kass & Raftery, 1995). Bayes factors (abbreviated here as BF) take the form of a likelihood ratio, measuring the evidence the data provide for one model over another, taking automatically into account both goodness-of-fit and model complexity (Lee & Wagenmakers, 2013, Chapter 7). For example, when $\mathrm{BF}_{34} = 5$ the observed data are five times more likely to have occurred under $\mathcal{S}_3$ than under $\mathcal{S}_4$; and when $\mathrm{BF}_{42} = .05$ the observed data are 20 times more likely to have occurred under $\mathcal{S}_2$ than under $\mathcal{S}_4$; moreover, because Bayes factors are transitive, $\mathrm{BF}_{32} = \mathrm{BF}_{34} \times \mathrm{BF}_{42} = 5 \times .05 = .25$, indicating that the observed data are four times more likely under $\mathcal{S}_2$ than under $\mathcal{S}_3$.

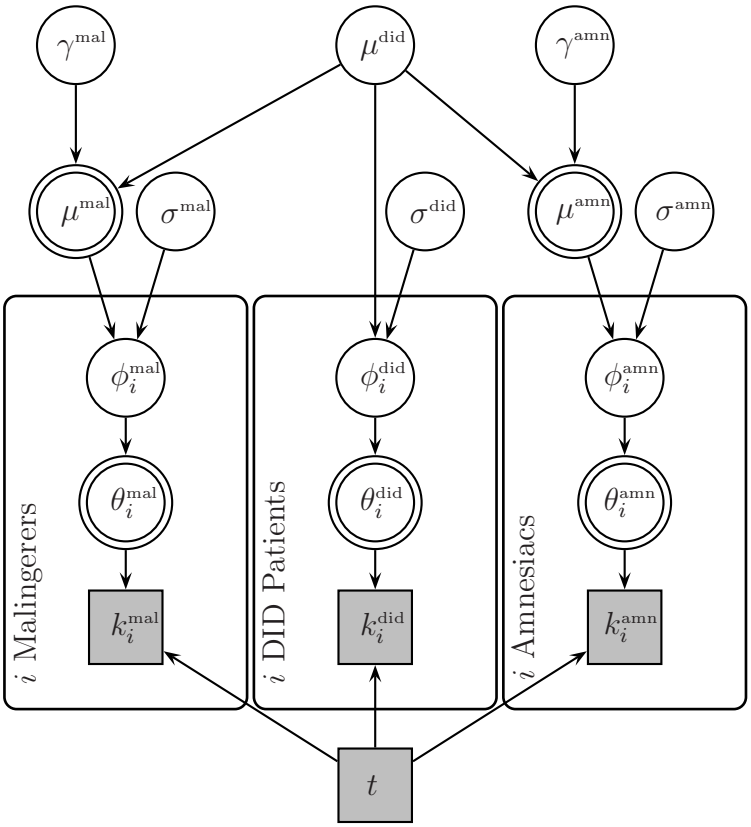

Figure 6. Graphical model framework for the Bayes factor comparisons of four different clusterings of the DID Patient, Malingering and Amnesiac groups.

The structures are achieved by setting

$$
\begin{aligned}
\mathcal{S}_1 \;&:\; \mu^{\text{mal}} = \mu^{\text{did}},\ \mu^{\text{amn}} = \mu^{\text{did}},\\
\mathcal{S}_2 \;&:\; \mu^{\text{mal}} = \mu^{\text{did}},\ \mu^{\text{amn}} = \mu^{\text{did}} + \gamma^{\text{amn}},\ \gamma^{\text{amn}} \sim \text{TruncGauss}_+\big(0,1\big),\\
\mathcal{S}_3 \;&:\; \mu^{\text{amn}} = \mu^{\text{did}},\ \mu^{\text{mal}} = \mu^{\text{did}} - \gamma^{\text{mal}},\ \gamma^{\text{mal}} \sim \text{TruncGauss}_+\big(0,1\big),\\
\mathcal{S}_4 \;&:\; \mu^{\text{amn}} = \mu^{\text{did}} + \gamma^{\text{amn}},\ \gamma^{\text{amn}} \sim \text{TruncGauss}_+\big(0,1\big),\\
&\quad\ \ \mu^{\text{mal}} = \mu^{\text{did}} - \gamma^{\text{mal}},\ \gamma^{\text{mal}} \sim \text{TruncGauss}_+\big(0,1\big).
\end{aligned}
$$

Because the different model structures follow a nested hierarchy, Bayes factors can be calculated easily using the Savage-Dickey density ratio test (for details see Dickey & Lientz, 1970; Lee & Wagenmakers, 2013; Wetzels, Grasman, & Wagenmak-

Table 2: Bayes factors for different model structures, calculated using Savage-Dickey density ratio tests using a truncated normal approximation to the posterior. See text for details.

	PD-1	PD-2	PD-3	PD-4	PD-5
BF_{41}	4715	4440	4427	3663	4663
BF_{42}	4.59	4.27	4.44	4.24	4.42
BF_{43}	3.04	2.90	3.02	2.98	2.84
BF_{32}	1.51	1.47	1.47	1.42	1.56

ers, 2010; Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010; computer code is available on the authors' websites). Table 2 lists the results.

The most likely model, $\mathcal{S}_4$, is the one in which each group has its own mean rate. As the first row of Table 2 shows, the evidence strongly favors this model over model $\mathcal{S}_1$, the model in which all groups have the same mean rate. In the two remaining models, the DID patients either have the same mean rate as the Malingerers (model $\mathcal{S}_2$), or they have the same mean rate as the Amnesiacs (model $\mathcal{S}_3$). Both these models are supported by the data slightly less than model $\mathcal{S}_4$, the model in which each group has its own mean rate. Finally, the bottom row of Table 2 shows that the models in which the DID patients are paired with the Malingerers is about as likely as the model in which the DID patients are paired with the Amnesiacs. Overall, the Bayes factors show that it is most likely each group has its own rate, although one cannot altogether rule out the possibility that the DID patients have the same mean rate as either the Malingerers or the Amnesiacs. However, the data provide almost no evidence to suggest which of these latter two hypotheses is true.

Conclusion

The three Bayesian analyses all address the basic research question—"does the recognition performance of the DID Patients resemble that of Malingerers or Amnesiacs?"—by asking different and complementary statistical questions. The first analysis showed that the estimated group accuracy of the DID Patients is about equidistant from the group accuracies of Malingerers and Amnesiacs. The second analysis showed that if DID Patients have to be classified as either Malingerers or Amnesiacs, they are evenly divided between the two. The third analysis shows that the most likely structure at the level of groups means is the structure in which DID Patients, Malingerers, and Amnesiacs each have their own unique mean.

Individually, none these analyses compellingly answer the research question at hand. But, in combination, they mount a strong argument that the research question cannot be clearly answered with the current data set. In fact, the results suggest that we can be relatively confident that we cannot be confident about whether DID Patients are more similar to Malingerers or Amnesiacs. By reasoning under uncertainty in a coherent manner, and by combining evidence from different analyses, we hope to have constructed an argument that even Rich Shiffrin will find convincing. Our prior beliefs, however, suggests that Rich is likely to suggest sensible analyses or alternative explanations that have not yet crossed our minds.

Acknowledgements

This work was supported in part by an ERC grant from the European Research Council.

References

Dickey, J. M., & Lientz, B. P. (1970). The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. *The Annals of Mathematical Statistics*, *41*, 214–226.

Gelman, A., & Stern, H. (2006). The difference between "significant" and "not significant" is not itself statistically significant. *The American Statistician*, *60*, 328–331.

Huntjens, R. J. C., Peters, M. L., Woertman, L., Bovenschen, L. M., Martin, R. C., & Postma, A. (2006). Inter–identity amnesia in dissociative identity disorder: A simulated memory impairment? *Psychological Medicine*, *36*, 857–863.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 377–395.

Klugkist, I. (2008). Encompassing prior based model selection for inequality constrained analysis of variance. In H. Hoijtink, I. Klugkist, & P. A. Boelen (Eds.), *Bayesian evaluation of informative hypotheses* (pp. 53–83). New York: Springer.

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology*, *55*, 1–7.

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge University Press.

Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). *The BUGS book: A practical introduction to Bayesian analysis.* Boca Raton (FL): Chapman & Hall/CRC.

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience*, *14*, 1105–1107.

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.* Vienna, Austria.

Rossell, D., Baladandayuthapani, V., & Johnson, V. E. (2008). Bayes factors based on test statistics under order restrictions. In H. Hoijtink, I. Klugkist, & P. A. Boelen (Eds.), *Bayesian evaluation of informative hypotheses* (p. 111-129). New York: Springer.

Shiffrin, R. M., Lee, M. D., Kim, W.-J., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, *32*, 1248–1284.

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*, 158–189.

Wetzels, R., Grasman, R. P. P. P., & Wagenmakers, E.-J. (2010). An encompassing prior generalization of the Savage–Dickey density ratio test. *Computational Statistics & Data Analysis*, *54*, 2094–2102.