

A Comparison of One-High-Threshold and Two-High-Threshold Multinomial Models of Source Monitoring

Mahesh Menon Todd S. Woodward
Riverview Hospital, Coquitlam, British Columbia, Canada

A data simulation study comparing the one-high-threshold (1HT) and two-high-threshold (2HT) multinomial models suggested that 2HT models are more likely to misestimate the underlying parameter values, due to inflation of some parameters (b and d), and deflation of others (D).

Key words: Multinomial modeling, source monitoring, data simulation

Introduction

Source monitoring and reality monitoring studies have proven to be extremely useful in understanding a variety of memory processes in the normal and clinical populations (Brebion, Gorman, Amador, Malaspina, & Sharif, 2002; Hoffman, 1997; Johnson, Hashtroudi, & Lindsay, 1993; Keefe, Arnold, Bayen, McEvoy, & Wilson, 2002; Lindsay, Johnson, & Kwon, 1991). Consider a simple source monitoring experiment with two sources: A and B. Single words are presented in a random fashion from the two sources, and the final recognition test consists of a mix of old A and B items along with new distracters N. The analysis of data from such a study typically examines item detection (the number of previously presented items that are correctly identified as being old), source recognition (the number of times the source of the item was correctly attributed), and

the false positive error rates (the number of new items that are incorrectly identified as being old).

However, traditional methods of analysis are unable to separate guessing biases and meta-cognitive response strategies from true item and source recognition. For example, when subjects notice that they are recognizing too few items from the (less memorable) external source, they tend to compensate by increasing the number of external-source guesses (Batchelder & Riefer, 1990). Therefore, in order to accurately measure externalizations, increases in strategic external guesses must be excluded. Similarly, to accurately estimate source recognition, it must be separable from both biases and strategic guessing. Multinomial modeling allows these distinct cognitive mechanisms can be disentangled. Multinomial modeling is a statistically sophisticated, yet simple method of separating item recognition, source recognition and response biases in discrimination tasks such as source monitoring, allowing estimation of guessing strategies and biases separately from source-discrimination processes.

Multinomial models attempt to explain discrete responses in a particular psychological paradigm by postulating latent cognitive processes that combine in different ways to determine the response category. The basic idea is that any given response category may occur as a consequence of one or more processing sequences, where each processing sequence is characterized by a series of successful or unsuccessful processing events. The processing

Mahesh Menon is also affiliated with the Schizophrenia PET Research Group, Centre for Addiction & Mental Health, Toronto, Canada. Email him at mahesh@cantab.net. Todd S. Woodward is also affiliated with the Department of Psychology, Simon Fraser University, Burnaby, British Columbia, Canada. Email him at toddswoodward@gmail.com.

sequences are represented in a tree structure (see Figure 1). The root (or initial node) represents the beginning of the processing sequence, the intermediate nodes represent stages involving a choice between two or more processing events, and the terminal nodes correspond to the observable response categories. The application of multinomial models to source monitoring has been reviewed in detail elsewhere (Batchelder & Riefer, 1999; Bayen, Murname, & Erdfelder, 1996).

One theoretical divide with implications for future research and interpretability of prior research findings lies in the selection of the basic model used for the analyses. Bayen et al. (1996) described three classes of models: the one-low-threshold (1LT), one-high-threshold (1HT) and two-high-threshold (2HT) models. The 1LT model has been used in some studies (Macmillan & Creelman, 1991), but the lack of a recognition bias parameter greatly weakens the LT model as a general purpose model of source monitoring (Bayen et al., 1996). The major theoretical debate therefore remains between the use of the 1HT and 2HT classes of models.

In 1HT models there is a single high threshold that divides the decision space into two discrete areas that correspond to detect as old and undetected. In 1HT models, it is assumed that only old items can cross the high threshold. If the threshold is crossed on presentation of a test item, the item is detected as old. If the threshold is not crossed, the item is said to be in an undetected state. All the new items and the old items that do not cross the threshold are categorized as old or new only on the basis of guessing. The probability with which an undetected item is guessed as being old is labeled b (see Figure 1).

In a 2HT model, there are two high thresholds that divide decision space into three discrete areas that correspond to detect as old, detect as new and undetected. It is assumed that only old items can cross the detect as old threshold, and only new items can cross the detect as new threshold. If either threshold is crossed on presentation of a test item, the item is detected as either old or new, depending on which threshold was crossed. If neither threshold is crossed, the item goes undetected, and is guessed to be new or old. As can be seen from

Figure 1, the 2HT source monitoring model can be constructed from the 1HT model by adding a parameter (labeled D_N) indicating the probability that a new item will be detected as new. Conversely, the 1HT model described earlier can be derived from the 2HT model by imposing the restriction $D_N = 0$. A 1HT model may thus always be regarded as a special case of a 2HT model, where the probability of crossing the second threshold is zero for all classes of items.

The nature of this debate centers around three issues. The primary theoretical issue is one of how an item is recognized as being old or new. In 2HT model space, the detection of an item as new requires it to cross a discrete threshold, which would determine it to be previously unseen. The 1HT model instead argues that the failure to cross the detection threshold for the item being old would imply that it remains unrecognized, and in the absence of a false positive recognition, the person would conclude that the item is new.

The second issue is an empirical one, stated as follows: for a given set of data, does including a D_N parameter affect the values of the other parameters (even though it is thought to represent a distinct, independent cognitive process), and if so, what is the nature of these changes?

A third issue lies in the interpretability of the parameters in the two classes of models. Specifically, in the 1HT and 2HT models (but not the 1LT model), the b parameter represents the probability of guessing that a word is a target item, when it has not actually crossed one of the thresholds. This parameter is reflected in two separate scenarios: (1) for previously seen words, it is an indicator of the tendency to guess that the word is old even when it has not been detected as being old, and (2) for new words, it is an indicator of the tendency to make false positive errors. It is assumed that these two scenarios are underpinned by the same cognitive process, and hence are assigned a common parameter. But how representative are the b parameters generated by the 1HT and 2HT model of this cognitive process?

The empirical question of how estimating a D_N parameter affects the values of the other parameters by way of a data simulation of a three-source source monitoring

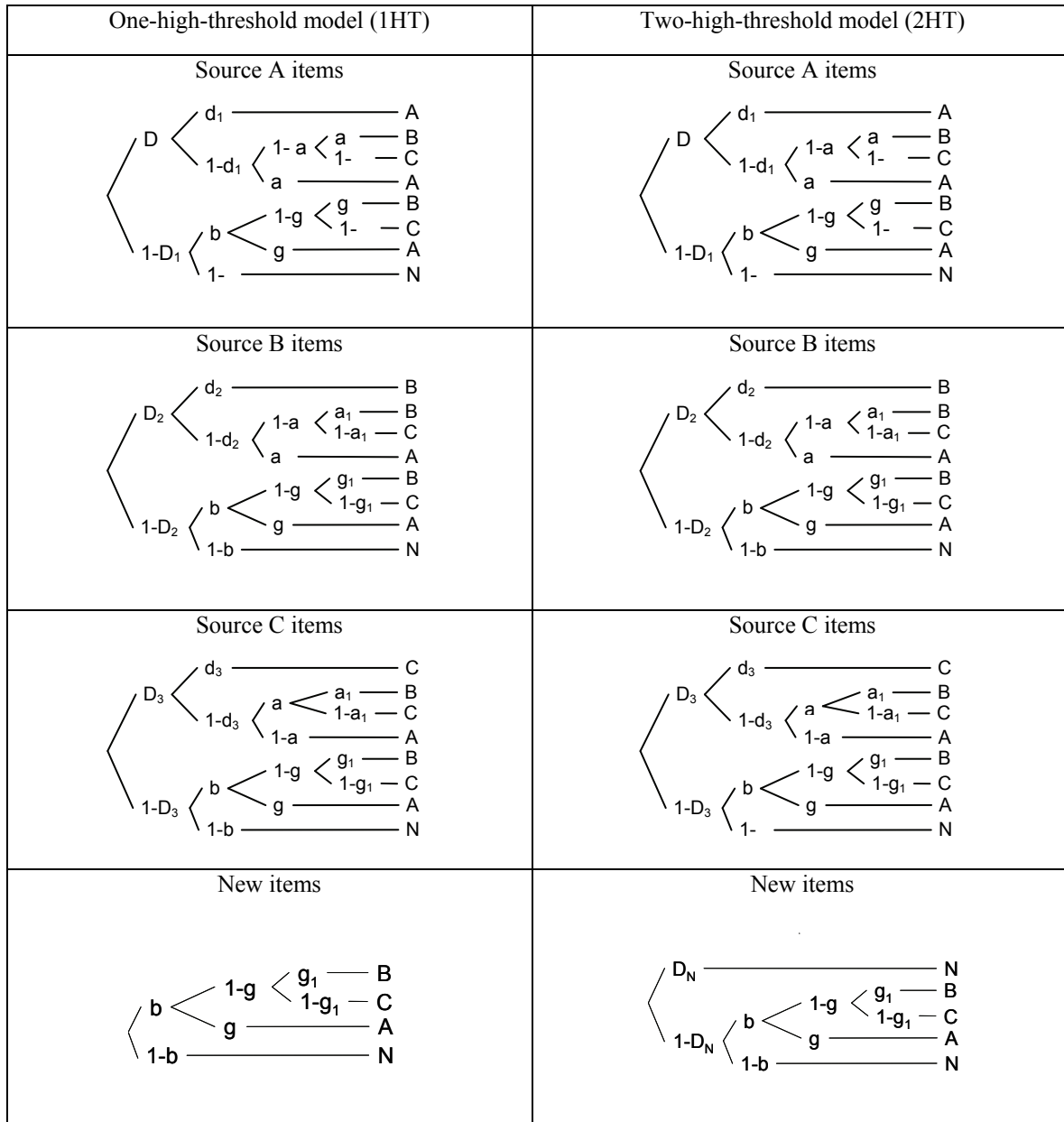


Figure 1 One- and two-high-threshold models of source monitoring.

Notes. A = Source A item; B = Source B item; C = Source C item; N = distracter item; Item recognition parameters: D_1 = probability of detecting an item from source A; D_2 = probability of detecting an item from source B; D_3 = probability of detecting an item from source C; D_N = probability of detecting that a distracter item is new; Source recognition parameters: d_1 = probability of correctly discriminating the source of an item from source A; d_2 = probability of correctly discriminating the source of an item from source B; d_3 = probability of correctly discriminating the source of an item from source C; Guessing biases: a = probability of guessing that a detected item is from source A; a_1 = probability of guessing that a detected item is from source B; b = probability of guessing that an undetected item is old; g = probability of guessing that an undetected item is from source A; g_1 = probability of guessing that an undetected item is from source B.

study was addressed. The three source model was used as it allows for greater flexibility with model specification, which is not limited by the available degrees of freedom (Keefe et al., 2002; Riefer, Hu, & Batchelder, 1994; Woodward et al., 2006). The data simulation results are also pertinent to the theoretical and interpretational issues mentioned above, and these issues are addressed in the Discussion section.

The simulation of the to-be-analyzed data involved generating frequency tables from tree models for which all parameter values were set, with the exception of D_N , which was varied between 0 and .90. Varying the D_N parameter in this fashion allowed us to simulate a situation where, as is assumed by the 1HT model, there is no cognitive process that is captured by D_N , as well as three other situations whereby detection of new items is the true state of affairs, to varying degrees.

Then, these frequency tables were analyzed using 1HT and various 2HT models. Performance was assessed by examining the inflation and deflation of estimated parameter values that occur when the 1HT and 2HT underlying assumptions did or did not match the true state of affairs (the true state of affairs being the state where the estimated D_N parameter value matched the generating D_N parameter value). In this fashion, the underlying assumptions (i.e., 1HT or 2HT generating parameter values) and the methods of analysis (i.e., 1HT or 2HT estimated parameter values) were completely crossed. Both the 1HT and 2HT models were expected to perform well when the method of analysis matched the underlying assumptions determined by the generating parameters, but inflations and deflations in parameter estimates were expected when a mismatch occurred between the method of analysis and the underlying assumptions. The conclusions are based on the net inflation or deflation of parameter values for the 1HT and 2HT models under these mismatch conditions.

Methodology

A series of simulated frequency tables of responses from a three-source source monitoring task were created, similar to one used by Keefe and colleagues (Keefe et al., 2002) and one used

by Woodward, Menon, and Whitman (in press). The frequency tables were created under the constraint of 100 old items from each of the three sources (labeled A, B and C), and 300 new items. A set of underlying parameter values (i.e., the generating parameters) were specified, and these were used to create sets of response frequencies, under the restrictions of 300 old and 300 new items, as mentioned above. The final probabilities for each source/response combination were computed by multiplying the generating parameters down the processing tree, and summing together the events that lead to each specific source/response combination. For example, for the 1HT and 2HT models shown in Figure 1, the probability of participants responding "A" given stimulus A can be arrived at by multiplying the following parameters:

$$P("A" | A) = D_1 * d_1 + D_1 * (1-d_1) * a + (1-D_1) * b * g \quad (1)$$

The simulated response frequencies were created by multiplying the final probabilities associated with each source/response combination (as specified in Figure 1, and exemplified in Equation 1) by the number of responses for that source. For example, multiplying the summed probability shown in Equation (1) by 100 generates the number of times participants responded "A" for stimuli from source A.

The generating parameter values were fixed across simulated sets of frequencies, with the exception of the D_N parameter (the item recognition parameter for new items), which was varied (0, .30, .60 and .90). This method of creating sets of frequencies is similar to the method used by Reifer and Batchelder (1991) in their Monte Carlo simulation. The sets of frequencies created when $D_N = 0$ represent patterns of frequencies that would be generated under the assumptions underlying 1HT model, whereas the sets of frequencies created when $D_N > 0$ represent patterns of frequencies that would be generated under the assumptions underlying various 2HT models. The frequency tables generated for the analyses are presented in Appendix A. Note that the only variation in the sets of frequencies generated using the various

values of D_N occurred in the number of false positive responses generated by the new words.

Then, the data were analyzed using an Excel spreadsheet specialized for multinomial modeling (Dodson, Prinzmetal, & Shimamura, 1998). In the analysis, the value of the D_N parameter was first fixed at the different values (0, .30, .60 and .90) to test the impact of varying degrees of mismatch between the underlying model assumptions (as determined by the generating parameter values) and the methods of analysis. Following this, the 2HT analysis was carried out allowing all parameters (including D_N) to vary freely. In the Excel spreadsheet, the optimum parameter values are found by allowing them to vary in an iterative fashion using the solver function. Starting values for probability estimates were .50, and the log likelihood ratio statistic G^2 was used to assess overall fit. G^2 asymptotically has a chi-square distribution, and the optimized solution is the one that minimizes the G^2 value (Riefer & Batchelder, 1988).

Results

The estimated parameter values are presented as a function of varying generating and estimating D_N parameter values in Table 1, and in Figures 2-6. From these results it is apparent that, within each set of response frequencies, varying the value of the D_N parameter (which in theory should only affect the recognition of new items) affected the value of the other parameters in a variety of important ways. As was expected, all models were very accurate when the method of analysis matched the underlying assumption (i.e., when the estimating D_N was equal to the generating D_N). However, in the case of mismatches, occasionally severe inflation and deflation of the underlying parameters occurred. Specifically, underestimation of D_N parameter resulted in overestimation of the other D parameters and underestimation of b and d parameters, whereas overestimation of D_N resulted in underestimation of D and overestimation of b and d parameters.

With respect to the comparative performance of 1HT and 2HT models, the most important pattern that can be derived from Table

1 and Figures 2-6 is that the consequences of overestimating D_N are far more severe than the consequences of underestimating D_N . As Figures 2-6 show, the changes in the generating parameter values had the least effect when estimated using the 1HT model (i.e., $D_N = 0$), while the variability of the estimated parameters increased greatly when a 2HT model is used, particularly with high D_N parameter values.

For example, consider the results based on a generating D_N parameter of .60 (see Table 1). An instructive comparison can be made between when D_N was underestimated by .30 and when D_N was overestimated by .30 for this condition. When the D_N parameter was underestimated by .30 (i.e., estimated $D_N = .30$), D_1 and D_2 increased to .70 (from .67), D_3 increased to .78 (from .75), and the d_1 and d_2 parameters decreased to .57 (from .60), a net change of .03 on all these parameters. This also resulted in a reduction in the value of the b parameter to .11 (from .20). However, in the comparison condition when D_N was overestimated by .30 (i.e., estimated $D_N = .90$), D_1 decreased to .40 (from .67), the d_{1-2} parameters increased to 1.0 (from .60, indicating perfect source recognition), a net change of .27 and .40, respectively. In addition, the b parameter increased to .59 (from .20).

Following the analysis where the estimating D_N parameter was fixed at various values, we also examined the results using an unconstrained 2HT model (i.e., all parameters, including the D_N parameter, were free to vary). The results are shown in Table 1 (bottom row). The unconstrained 2HT model did not retrieve the generating parameters in any of the lower D_N conditions (i.e., generating $D_N = 0, .30$ or .60), despite a perfect fit (i.e., very low G^2 values). Instead the model tended to estimate D_N values that were higher than the generating D_N values, resulting in a corresponding elevation of the b and d parameter values, and a reduction in the D parameter values.

Both 1HT and 2HT models estimated the guessing parameters a , a_1 , g and g_1 accurately, and their values did not change with changes in the value of the estimating D_N parameter; therefore, they are not presented here.

Table 1. Estimated Parameter Values Presented as a Function of Varying Generating and Estimating D_N Parameter Values

Generating D_N Parameter													
$D_N = 0$ (1 HT) FP = .20							$D_N = .30$ FP = .14						
Est D_N	D_{1-2}	D_3	d_{1-2}	d_3	b	Fit	Est D_N	D_{1-2}	D_3	d_{1-2}	d_3	b	Fit
$D_N = 0$.67	.75	.60	.59	.20	0 (.99)	$D_N = 0$.69	.77	.58	.58	.14	0 (.99)
$D_N = .30$.63	.72	.64	.61	.29	0 (.99)	$D_N = .30$.67	.75	.60	.59	.20	0 (.99)
$D_N = .60$.47	.60	.85	.74	.50	0 (.98)	$D_N = .60$.59	.69	.68	.64	.36	0 (.99)
$D_N = .90$.37	.44	1.0	1.0	.68	62.8 (.0)	$D_N = .90$.38	.45	1.0	1.0	.65	26.3 (.0)
$D_N = .53$.54	.66	.74	.68	.43	0 (.99)	$D_N = .61$.58	.69	.69	.65	.37	0 (.99)

Generating D_N Parameter													
$D_N = .60$ FP = .08							$D_N = .90$ FP = .02						
Est D_N	D_{1-2}	D_3	d_{1-2}	d_3	b	Fit	Est D_N	D_{1-2}	D_3	d_{1-2}	d_3	b	Fit
$D_N = 0$.71	.78	.57	.56	.08	0 (.99)	$D_N = 0$.73	.80	.55	.55	.02	0 (.99)
$D_N = .30$.70	.78	.57	.57	.11	0 (.99)	$D_N = .30$.73	.80	.55	.56	.03	0 (.99)
$D_N = .60$.67	.75	.60	.59	.20	0 (.99)	$D_N = .60$.72	.79	.56	.56	.05	0 (.99)
$D_N = .90$.40	.52	1.0	.74	.59	2.6 (.10)	$D_N = .90$.67	.75	.60	.58	.20	0 (.99)
$D_N = .73$.63	.72	.65	.62	.30	0 (.99)	$D_N = .91$.67	.75	.61	.59	.21	0 (.99)

Note. The generating probabilities used were: $D_1/D_2 = .67$, $D_3 = .75$, $d_1/d_2 = .60$, $d_3 = .60$, $a = .60$, $b = .20$, $g = .60$, $a_1 = .50$, $g_1 = .50$. FP refers to false positives, and Fit refers to Chi-square values of the final model (p values are bracketed, such that $p < .05$ indicates poor fit). When estimating, all parameters were free to vary with the exception of D_N , which was fixed (Est D_N). The bottom row displays the results when the D_N parameter was also allowed to vary. The models generated guessing parameters (a , a_1 , g , g_1) that were identical to the generating probabilities and did not vary across the different simulations, and are therefore not listed in the above table.

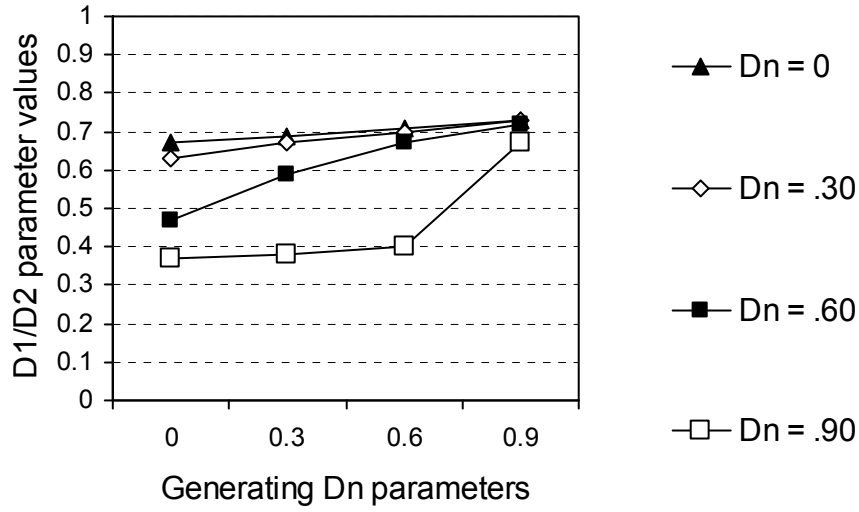


Figure 2 Estimated D_1/D_2 parameter values plotted as a function of generating and estimated D_N parameters (True $D_1/D_2 = .67$).

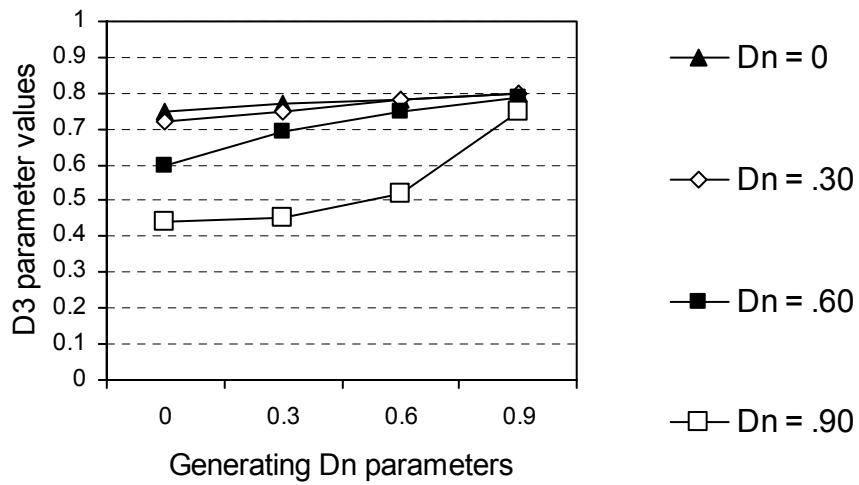


Figure 3. Estimated D_3 parameter values plotted as a function of generating and estimated D_N parameters (True $D_3 = .75$).

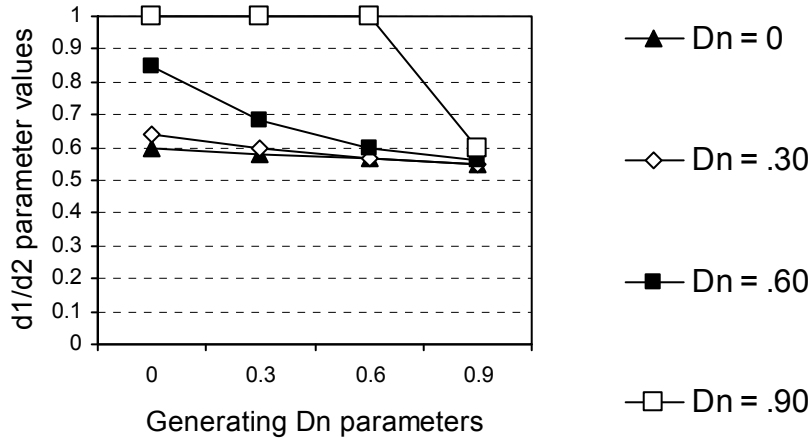


Figure 4 Estimated d_1/d_2 parameter values plotted as a function of generating and estimated D_N parameters (True $d_1/d_2 = .60$).

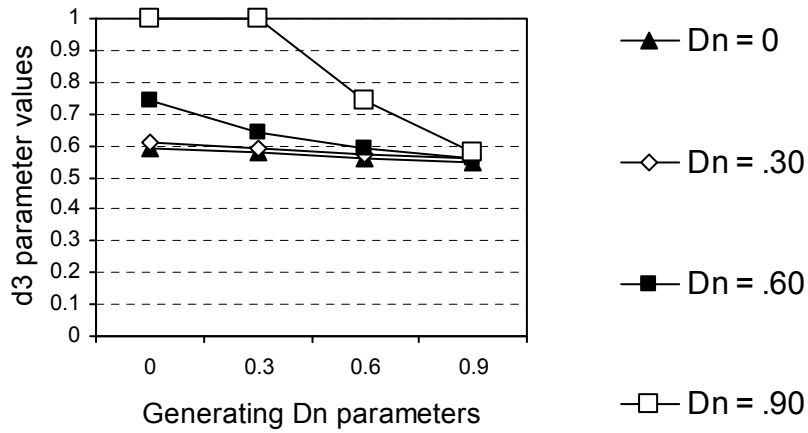


Figure 5. Estimated d_3 parameter values plotted as a function of generating and estimated D_N parameters (True $d_3 = .60$)

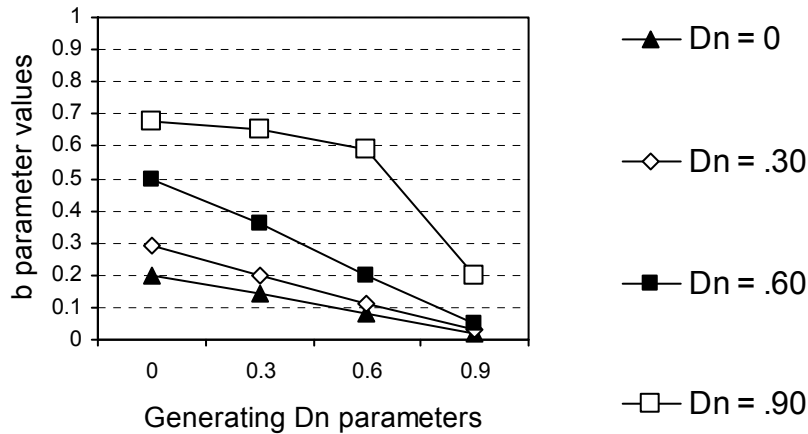


Figure 6. Estimated b parameter values plotted as a function of generating and estimated D_N parameters (True $b = .20$).

Conclusion

This simulation was designed to compare the 1HT and 2HT approaches to multinomial modeling on their ability to accurately estimate underlying parameter values under a number of experimental conditions. The results suggest that the 1HT model is a more conservative choice for the analysis of data from source monitoring tasks, because the 2HT models are more likely to misestimate the underlying parameters. Specifically, the simulation showed that use of the 2HT models lead to an artificial inflation of the b parameter (probability of guessing that an item not detected as old was actually seen before), which in turn caused inflation of d (source recognition), and deflation of D (item recognition) parameter values. The 1HT model showed less variability and gave parameter estimates that were closer to the underlying parameter values, even when the underlying assumptions were those held by the 2HT perspective. As the simulation shows, the unconstrained 2HT solution (where the D_N parameter value was not fixed - Table 1 bottom row) typically produced a D_N parameter value

that was in excess of the true generating D_N value, which in turn affected many of the other parameters as outlined above. With respect to a behavioral interpretation, the 2HT models produce parameter values that suggest artificially reduced item recognition for old items, increased guessing of undetected old items, and increased source recognition compared to the 1HT model. For instance, with the high estimated D_N parameter values, the 2HT models suggest a state with perfect source recognition even though item recognition is occurring much less frequently, which is counterintuitive, given that item recognition is generally regarded as a less demanding process than source recognition.

With respect to the theoretical issue mentioned in the Introduction, whether or not the recognition of new items occurs via the new items crossing a definite threshold (as suggested by the 2HT model) or simply by not being recognized as being old (as suggested by the 1HT model) remains open to debate. What the simulation results show is that, given that these theoretical issues cannot currently be resolved, the 1HT model is a more conservative choice for

the analysis of data from source monitoring tasks, because the 2HT models are more sensitive to possible mis-estimation of the underlying parameters.

With respect to interpretability of the b parameter (the probability of guessing that an undetected item is actually a target item), as can be seen in Table 1, the results show that only the 1HT model (i.e., $D_N = 0$) produced a b parameter value equal to the false positive error rate. In the 2HT model, increasing values of D_N resulted in b parameter values that were far higher than the proportion of false positives. As mentioned earlier, the parameter is an indicator of two separate scenarios. For previously seen words, it is an indicator of the tendency to guess that the word is old even when it has not been detected as being old, and for new words, it is the tendency to make false positive errors (i.e., detect them as being old). However, the crucial difference between these two processes is that the tendency to make false positive errors can be estimated (by observing raw frequencies of the number of false positive recognition errors), while the same cognitive process for old words (making fortuitous false positive recognitions, which in turn elevates the hit rate), remains a hidden process.

If both these patterns of false recognition are driven by the same cognitive processes, then the b parameter should reflect both in equal measure. The use of the 1HT model, which generates the value of the b parameter corresponding to the proportion of false positive error rates, has an intuitive appeal from this perspective. The 2HT model is likely to produce a b parameter that is inflated, possibly leading to the false conclusion that much of the recognition reflected by the raw frequencies occurs through fortuitous guessing. Moreover, an increase in b causes corresponding decreases and increases in the D and d parameters, respectively, leading to the false conclusion that recognition has decreased and source discrimination has increased, respectively.

In an important study that suggested the superiority of the 2HT model over the 1HT model in source monitoring, Bayen et al. (1996) studied the impact of increased distracter similarity on the ability to carry out item

detection and source recognition. On the assumption that increased distracter similarity (but not increased source similarity) would decrease item detection, they analyzed their results using 1HT and 2HT models. They found that D parameter values decreased with increasing distracter similarity only when the 2HT model was used, and argued on these grounds that the 2HT is more sensitive to changes in item detection, making it a superior model of source monitoring. However, as Bayen et al. (1996) pointed out, increasing distracter similarity only serves to increase the number of false alarms and not the hit rate (p 205, Appendix C pg 215), casting doubt on their assumption that increasing distracter similarity should decrease item detection. Moreover, the simulation results demonstrate that use of the 2HT model is likely to lead to artificial decreases in the recognition parameter D , suggesting that Bayen et al.'s (1996) recommendations may have been based on an artifact of the 2HT model, as opposed to the purported superiority of the 2HT model in detecting an experimental-manipulation-induced true decrease in recognition.

There are two reasons to suggest that the recognition parameters may have been affected by the model selected rather than the experimental manipulation. First, the Bayen et al. (1996) models used one D parameter for both old and new items (i.e. $D_N = D_1 = D_2$), and as a result, the increase in false positives, which should cause a decrease in D_N but not in D_1 or D_2 , affected all three of these parameters. Secondly, due to the reciprocal relationship between b and D in the 2HT model (as can be seen from Figures 2, 3 and 6), an increase in the b parameter (which is often well in excess of the true proportion of false alarms in the experimental data) caused a reduction in the D parameter. That is to say, use of the 2HT model can lead to apparent decreases in item detection as an artifact of increases in false positives, even in the absence of an experimental manipulation affecting item detection.

In the aforementioned Bayen et al. (1996) study, the 1HT model, on the other hand, showed a slight decrease in the D parameter due to an increase in false alarms, and a significant decrease in the D parameter only when the hit

rate decreased. This pattern of results would be expected from a model that was appropriately reflecting factors affecting false alarms independently from those affecting item detection.

Bayen et al. (1996) computed decreases in the signal detection measure d' for item detection. They showed decreases in d' with increasing source similarity, and, based on this measure, concluded that their experimental manipulations had affected the true recognition rate. Batchelder, Riefer and Hu (1994) and Thomas and Olzak (1992) explicitly pointed out that d' can be used as a means to compare item recognition only when false alarm rates are comparable across conditions. For a paradigm such as that used by Bayen et al. (1996), where the manipulation specifically affected the false alarm rates, d' is expected to underestimate item detection as the false alarm rate increases. It therefore seems likely that the fundamental assumption of their experimental manipulation (i.e., that increasing distracter similarity causes decreases in item detection) is debatable, and that the decreases in recognition parameters that they observed when employing the 2HT model may have instead reflected an artifact of the 2HT model.

The data simulation indicates that the 1HT model generally provides more accurate estimates of the underlying parameter values than the 2HT model, and is more robust to variation in the generating D_N parameters. In addition, with respect to interpretation, the 1HT model produces a b parameter value that accurately reflects the true proportion of false positives and the 2HT model inflates the b parameter value, which in turn spuriously reduces the item detection parameters and inflates the source recognition parameters. In the light of these considerations, use of the 1HT model over the 2HT model is recommended in the analysis of data from source monitoring studies.

References

- Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, *97*, 548-564.
- Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*, 57-86.
- Batchelder, W. H., Riefer, D. M., & Hu, X. G. (1994). Measuring memory factors in source monitoring - reply to kinchla. *Psychological Review*, *101*, 172-176.
- Bayen, U. J., Murname, K., & Erdfelder, E. (1996). Source monitoring, item detection and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*, 197-215.
- Brebion, G., Gorman, J. M., Amador, X., Malaspina, D., & Sharif, Z. (2002). Source monitoring impairments in schizophrenia: Characterisation and associations with positive and negative symptomatology. *Psychiatry Research*, *112*(1), 27-39.
- Dodson, C. S., Prinzmetal, W., & Shimamura, A. P. (1998). Using excel to estimate parameters from observed data: An example from source memory data. *Behavior Research Methods, Instruments, & Computers*, *30*, 517-526.
- Hoffman, H. G. (1997). Role of memory strength in reality monitoring decisions: Evidence from source attribution biases. *Journal of experimental Psychology: Learning, Memory and Cognition*, *23*(2), 371-383.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3-28.
- Keefe, R. S. E., Arnold, M. C., Bayen, U. J., McEvoy, J. P., & Wilson, W. H. (2002). Source-monitoring deficits for self-generated stimuli in schizophrenia: Multinomial modeling of data from three sources. *Schizophrenia Research*, *57*(1), 51-67.

Lindsay, D. S., Johnson, M. K., & Kwon, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology*, 52, 297-318.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge, England: Cambridge University Press.

Riefer, D. M., & Batchelder, W. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, 95(3), 318-339.

Riefer, D. M., & Batchelder, W. (1991). Statistical inference for multinomial processing tree models. In J.-P. Doignon & J.-C. Falmange (Eds.), *Mathematical psychology: Current developments*. New York: Springer-Verlag.

Riefer, D. M., Hu, X. G., & Batchelder, W. H. (1994). Response strategies in source monitoring. *Journal of Experimental Psychology-Learning Memory and Cognition*, 20(3), 680-693.

Thomas, J. P., & Olzak, L.A. (1992). Simultaneous detection and identification. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 253-277). Hillsdale, NJ: Erlbaum.

Woodward, T. S., Menon, M., & Whitman, J. C. (in press). Multinomial modelling of source monitoring biases associated with hallucinations and delusions in schizophrenia.

Woodward TS, Menon M, Hu X, Keefe RSE. (2006). Optimization of a multinomial model for investigations of hallucinations and delusions with source monitoring. *Schizophrenia Research*, 85, 106- 112.

Appendix A

Table of frequencies generated with varying 2HT assumptions

Source	Response			
	Experimenter	Computer	Self	New
$D_N = 0$				
Experimenter	50 (.50)	10 (.10)	13 (.13)	26 (.26)
Computer	10 (.10)	50 (.50)	13 (.13)	26 (.26)
Self	11 (.11)	11 (.11)	58 (.58)	20 (.20)
New	18 (.06)	18 (.06)	24 (.08)	240 (.80)
$D_N = .30$				
Experimenter	50 (.50)	10 (.10)	13 (.13)	26 (.26)
Computer	10 (.10)	50 (.50)	13 (.13)	26 (.26)
Self	11 (.11)	11 (.11)	58 (.58)	20 (.20)
New	13 (.04)	13 (.04)	17 (.06)	258 (.86)
$D_N = .60$				
Experimenter	50 (.50)	10 (.10)	13 (.13)	26 (.26)
Computer	10 (.10)	50 (.50)	13 (.13)	26 (.26)
Self	11 (.11)	11 (.11)	58 (.58)	20 (.20)
New	7 (.02)	7 (.02)	10 (.03)	276 (.92)
$D_N = .90$				
Experimenter	50 (.50)	10 (.10)	13 (.13)	26 (.26)
Computer	10 (.10)	50 (.50)	13 (.13)	26 (.26)
Self	11 (.11)	11 (.11)	58 (.58)	20 (.20)
New	2 (.007)	2 (.007)	2 (.007)	294 (.98)

Note. Row percentages are presented in brackets. Correct responses are in bold.