ORIGINAL PAPER

Considerations for analysis of source monitoring data when investigating hallucinations in schizophrenia research

Todd S. Woodward · Mahesh Menon

Abstract Self/other (i.e., internal/external) source monitoring is one of the leading paradigms for the study of hallucinations in schizophrenia. The cognitive processes that underlie hallucinations are theorized to transform self-generated (internal) cognitive events into other-generated (external) cognitive events. These proposed cognitive operations also appear to play a role in producing analogous types of errors in self/other source monitoring, namely a memory bias whereby recalled material that was self-generated is misremembered as other-generated, referred to as an *externalization bias*. Externalization biases are more frequent in groups of hallucinating schizophrenia patients than in other groups. One source of measurement error that is inherent in the study of the externalization bias is that, even for never-previously viewed items, there is a tendency to guess an external source under conditions of uncertainty. If such guessing takes place in response to self-generated but forgotten items, these guesses will be summed along with true externalization biases in the frequency count of externalizations, producing measurement error. Multinomial modeling is a statistical technique that has been used to estimate the influence of external-source guessing in order to separate it from true externalization bias estimates. However, a number of challenges related to model choice and model validation are involved, and these challenges may render multinomial modeling impractical. We instead recommend analysis of covariance (ANCOVA), or difference score methodology, as an appropriate method for partialling external-source guessing rates (external-source false positives) out of externalization bias rates.

T. S. Woodward
Department of Psychiatry, University of British Columbia,
Vancouver, Canada

T. S. Woodward (✉)
Department of Research, BC Mental Health & Addictions
Research Institute, Room 116, 3rd Floor,
938 W. 28th Avenue, Vancouver, BC V5Z 4H4, Canada
e-mail: toddswoodward@gmail.com

M. Menon
Schizophrenia Program and PET Centre,
Centre for Addiction & Mental Health, Toronto, Canada

M. Menon
Department of Psychiatry,
University of Toronto, Toronto, Canada

Introduction

The source memory paradigm has proven to be helpful in our understanding of hallucinations in schizophrenia [1–3]. In a typical source monitoring experiment designed for the study of hallucinations, words are encountered that originate from three sources: experimenter, computer, and self. The recognition test involves encountering old items from all three sources along with new distracters, and the subject is required to indicate whether the encountered word was old or new, and if old, from which of the three sources it originated. A matrix of response frequencies from such an experiment is presented in Table 1 (these data are taken from [1]). The analysis of data from such a study typically examines item detection (previously presented items that are correctly identified as old), source recognition (recognized items for which the source was correctly

Table 1 Response frequencies for 15 hallucinating schizophrenia patients as a function of source and response, on a source monitoring test with 25 each of self-, computer- and experimenter-generated items and 25 new items (resulting in 100 presented items)

Source	Response			
	Experimenter	Computer	Self	New
Experimenter	**181 (48)***	50 (13)#*	9 (2)§	135 (36)∀
Computer	98 (26)#*	**125 (33)***	9 (2)§	143 (38)∀
Self	46 (12)†*	81 (22)†*	**204 (54)**	44 (12)∀
New	33 (9)⊥*	28 (7)⊥*	3 (1)⊥	**311 (83)**

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

† Externalizations

§ Internalizations

Confusion of external sources

⊥ False positives

∀ False negatives

* Counts that would be inflated by guessing either of the two external sources

discriminated), and the false-positive error rates (the number of new items that are incorrectly identified as being old); however, none of these are the parameters that are of primary interest for the study of hallucinations. What is of primary interest is a memory bias whereby recalled material that was self-generated is misremembered as other-generated, and this is referred to as an *externalization bias*. Given that auditory hallucinations can be conceptualized as the assignment of internally generated mental episodes to an external source, an association between hallucinations and an externalization bias would suggest overlapping cognitive operations. An association between hallucinations and an externalization bias has been reported for schizophrenia patients [1, 2, 4, 5] and for healthy subjects who report hearing voices [6].

An interpretational challenge associated with source monitoring data is to separate the following two types of externalization errors: (1) when an internally generated event has been transformed to an external event by the same cognitive processes that produce hallucinations and (2) when the subjects simply guessed that the source was external for an internally generated event. The latter guessing events may be common. 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 [7]. This can occur on any type of trial whether or not the presented item was new, internally generated, or externally generated. The cognitive processes that generate these types of guessing responses are presumably not the cognitive processes that are involved in hallucinations and are therefore not the processes of interest when studying hallucinations. When these guessing processes take place in response to internally generated but forgotten items, these errors will be summed with true externalization biases in the count of externalizations. Therefore, in order to accurately quantify true externalizations that are of interest to the study of hallucinations, increases in these strategic "external" guesses must be excluded.

Multinomial modeling is a statistical technique that is recognized as showing potential for allowing external-source guessing strategies to be extracted out of true externalization bias estimates. 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 sequences are represented in a tree structure (see Figs. 1 and 2). 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 (e.g., [8]).

The application of multinomial modeling to the study of the positive symptoms of schizophrenia has been carried out under the motivation of separating guessing from true biases. In one study [9], a model was used that was adapted from the study of healthy cognitive processes in cognitive psychology (similar to that displayed in Fig. 1), and in another [10], a new model was developed that overcame the limitations of the established model for the study of hallucinations in schizophrenia (similar to that displayed in Fig. 2). However, both of these models are subject to limitations that will be reviewed below. We begin by presenting models developed within the field of cognitive psychology, discuss their limitations with respect to the study of hallucinations, and then go over the challenges involved in creating a new model that would be helpful for meeting the objective of investigating the cognitive underpinnings of hallucinations (i.e., measuring externalization biases free of the influence of guessing).

Limitations of current multinomial models

Established multinomial models of source monitoring (e.g., [9, 11]) separate guessing patterns in response to new or not recognized items (parameters b and g in Fig. 1) from "guessing" patterns in response to recognized but not accurately source discriminated items (parameters a and e in Fig. 1). However, for the study of hallucinations, it is
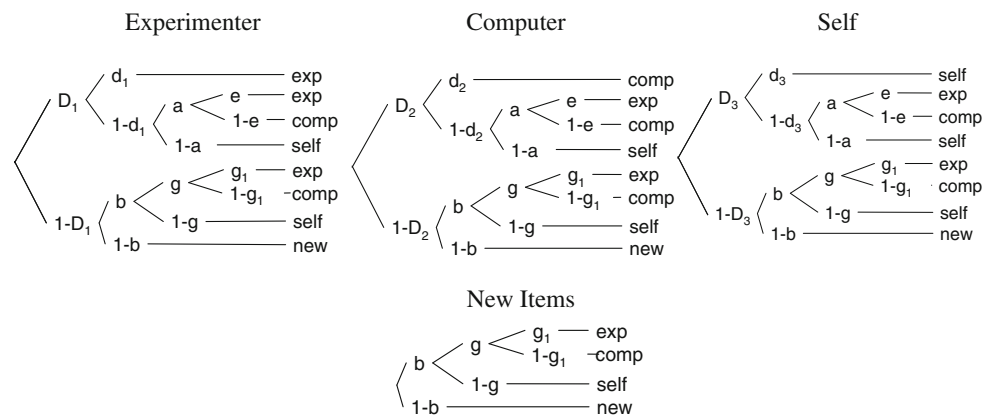
Fig. 1 Tree diagrams for a traditional three-source multinomial model (e.g., [9], with separate trees for self-, experimenter-, computer-generated and new items. D_1 = probability of recognizing experimenter-generated items as old; D_2 = probability of recognizing computer-generated items as old; D_3 = probability of recognizing self-generated items as old; d_1 = probability of discriminating the source of recognized experimenter-generated items; d_2 = probability of discriminating the source of recognized computer-generated items; d_3 = probability of discriminating the source of recognized self-generated items; a = probability of responding with an external source to a recognized-but-not-discriminated item; e = probability of responding "experimenter" to a recognized-but-not-discriminated item that is attributed to an external source; b = probability of responding 'old' to unrecognized items; g = probability of guessing that a unrecognized item belongs to an external source; g_1 = probability of guessing that a unrecognized item originated with the experimenter
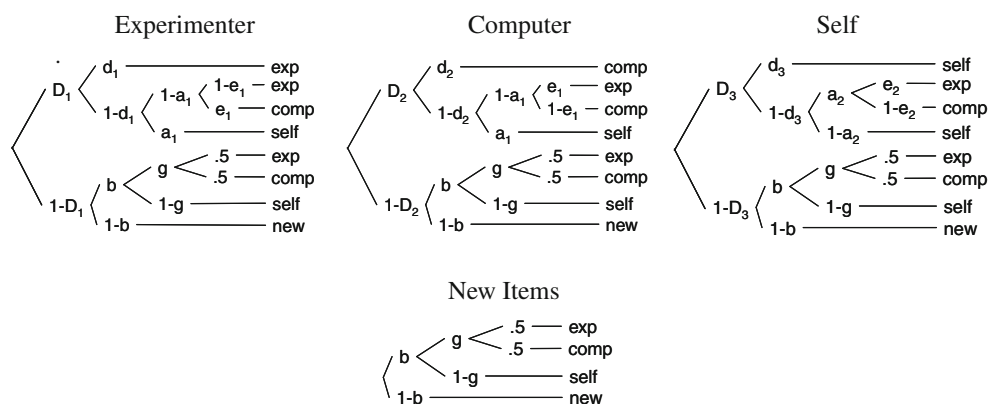

Fig. 2 Tree diagrams for the alternative three-source multinomial model (e.g., [10]), with separate trees for self-, experimenter-, computer-generated and new items. D_1 = probability of recognizing experimenter-generated items as old; D_2 = probability of recognizing computer-generated items as old; D_3 = probability of recognizing self-generated items as old; d_1 = probability of discriminating the source of recognized experimenter-generated items; d_2 = probability of discriminating the source of recognized computer-generated items; d_3 = probability of discriminating the source of recognized self-generated items; a_1 = probability of internalization when externally generated items are recognized but not discriminated; a_2 = probability of externalization when self-generated items are recognized but not discriminated; e_1 = probability of confusing the two external sources when externally generated items are recognized but not discriminated; e_2 = probability of attributing a self-generated item to the experimenter when self-generated items are recognized but not discriminated; b = probability of responding 'old' to unrecognized items; g = probability of guessing that a unrecognized item belongs to an external source. The key difference between this model and that presented in Fig. 1 is that the a parameter from Fig. 1 is split into the source-dependent parameters a_1 and a_2 in Fig. 2, with the former being estimated for external source (computer or experimenter) misattributions to self (internalizations), and the latter being estimated for self source misattributions to an external source (computer or experimenter; externalizations). Since it is the externalizations that are of interest for hallucinations, and these cannot be estimated using the model depicted in Fig. 1, the model depicted in Fig. 2 is desired for the study of hallucinations (for further details, see Woodward et al. [10]). However, this is a saturated model, and has not been validated through cognitive experimentation

necessary to estimate *separate* error parameters for recognized-but-not-discriminated items originating from the *external* source and those originating from the *internal* source. As was discussed above, it is the externalization errors on the internally generated items that are of primary interest to the study of hallucinations. However, the established multinomial models do not allow separate estimates of true internalization and externalization biases, because parameter estimates of biases *do not vary with the originating source* as would be required for the study of externalizations for hallucinations. That is to say, in the established models, unlike the parameters estimated for

recognition and source discrimination (D_{1-3} and d_{1-3}), biases affecting recognized-but-not-discriminated items are equated for items with internal- and external-source origins (referred to as *a* and *e* in Fig. 1 below).

An alternative model is presented below (Fig. 2), and an analysis based on a related model has been previously published by our group [10]. As can be observed by comparing Fig. 1 and 2 below, the relevant difference between the models is based on estimating parameters that vary with *source*. The *a* parameter from Fig. 1 is split into the source-dependent parameters a_1 and a_2 in Fig. 2, with the former being estimated for external-source (computer or experimenter) misattributions to self (i.e., internalizations) and the latter being estimated for self-source misattributions to an external source (computer or experimenter; i.e., externalizations). Since it is the externalizations that are of interest for the study of hallucinations and these cannot be estimated using the model depicted in Fig. 1, the model depicted in Fig. 2 would be superior for the study of hallucinations (for further details, see [10]). However, although conceptually this model meets the goals of research on hallucinations, there are challenges involved with using such a model as a standard protocol.

Challenges in developing a multinomial model for the study of hallucinations

A multinomial model that meets the requirements for the study of the externalization bias in hallucinations can be derived. One such model is presented in Fig. 2, and related model is presented in our past work [10]. Although not obvious at first glance, the model presented in Fig. 2 meets restrictive criteria (explained below). These criteria how they must be met and the factors that must be considered when making these decisions are outlined below.

Number of parameters that can be estimated

The number of parameters that can be estimated using multinomial modeling is limited by the total degrees of freedom. The total number of degrees of freedom is equal to the number of values in the frequency table that comprises the source of data for multinomial modeling, with one degree of freedom subtracted for each row in the frequency table. In the frequency table listed in Table 1, there are 12 *df* (16 values—4 rows = 12 *df*); therefore, only 12 separate parameters can be estimated. For example, 11 parameters were estimated to fit Fig. 1 and 12 parameters to fit Fig. 2 (the latter resulting in a saturated model). Thus, due to this limitation on the number of estimable parameters, even when using a saturated model as we did in Fig. 2, separate *a* parameters could not be estimated for each source separately (i.e., only a_1 and a_2 could be

estimated, not a_3, which would produce an estimate for each source).

Model fit

Even when a model is developed that estimates a restricted number of parameters (as is dictated by the available degrees of freedom), it remains to be determined whether or not the model fits. A fit statistic referred to as G^2 is typically used for this purpose. The log likelihood ratio statistic G^2 asymptotically has a χ^2 distribution [12], and significance of G^2 indicates that the proposed model does not fit the data. One complication is that G^2 cannot be computed on a saturated model such as that presented in Fig. 2, because there are no available degrees of freedom, and further restrictions must be applied in order to test the fit of the model. For example, if theoretically appropriate and practical, the D_1 parameter may be equated to the D_2 parameter, saving a degree of freedom and allowing model fit to tested with G^2 and one degree of freedom. If the model does not fit at any stage of the process, including after employing the latter strategy for saving a degree of freedom, it must be redesigned until acceptable fit is achieved.

1-HT vs. 2-HT model choice

One crucial theoretical divide within the multinomial modeling literature lies in the selection of the general model type used for the analyses. The major debate is between the use of the one-high-threshold (1HT) and two-high-threshold (2HT) classes of models, which differ in that the latter estimates a parameter for the detection of new items where the former does not (i.e., the 2HT model suggests that detection of items as 'new' involves a unique set of cognitive processes, while the 1HT model simply assumes that an item that is not recognized as old will be classified as 'new'). Some authors have recommended the 2HT model [11]; however, in our past work [13], we have shown that the 2HT model produced *b* values substantially higher than the false-positive rate and *D* values substantially lower than the recognition rates, respectively, raising concerns about the appropriateness of the 2HT model. Moreover, use of the 2HT model costs 1 *df* (it requires estimation of a D_4 parameter reflecting the number of new items "recognized" as new) and therefore further restricts the number of parameters that can be estimated. This decision is crucial to model development, but clear guidelines for how to make this choice are not readily available.

Validation challenges

Adjustments to traditional multinomial models may require extensive validation on healthy subjects prior to widespread

acceptance [11, 14, 15]. For example, if a cognitive process such as externalization is proposed in a new model, as it is in Fig. 2, cognitive experiments that manipulate the rate of externalizations (e.g., manipulating the perceptual vividness of self-generated material) must be carried out to prove that the proposed estimate is actually indexing some real cognitive process that varies with the experimental manipulation prior to utilizing the model. Bayen et al. [11] concluded that they had validated their presented model by presenting data, suggesting that the manipulation of difficulty in item detection was reflected in item detection and not in source discrimination parameters and that manipulation of difficulty in source discrimination was reflected in source discrimination and not in item detection parameters. However, a data simulation approach used by our group suggests that instead the recognition parameters (D values) may have been affected by the nature of the specific 2HT model fit rather than the experimental manipulation [13].

Assumption of response distribution equivalence

A general assumption of multinomial models is that the distribution of responses must be identical for every subject [16], which may be very difficult to meet for heterogeneous patient samples. This assumption is of particular importance when correlational analyses are to be carried out on the derived parameters, because parameters can vary widely if the model does not fit for each individual (with respect to G^2). Importantly, and even more restrictively, when making group comparisons on a given model, the model must fit for each individual group before between-group comparisons can be carried out.

Using ANCOVA instead of multinomial modeling for schizophrenia research

The overall goal of multinomial modeling is to eliminate the contribution of guesses from the parameter estimates, but due to the restrictions in the number of parameters that can be estimated, fit, model choice, and validation concerns (as discussed above), we cannot typically estimate source-specific biases or source-specific guessing estimates. As mentioned previously, the b parameter in multinomial models (the probability of guessing that an undetected item is actually a target item) is conceptually, and in some cases empirically (1HT models), equal to the false-positive error rate. That is to say, for some models, b need not be estimated, because this is already available in the form of the rate of false positives.

Since b is equal to the false-positive rate and is used to eliminate the contribution of guesses from the parameter estimates in multinomial modeling, it is informative to consider the common practice in memory experiments (and

in other standard clinical neuropsychological contexts of memory assessment), whereby a "corrected recognition" index is computed. This involves subtracting (i.e., eliminating) the false-positive rate from the recognition rate in order to provide a measure of recognition that is not inflated by guesses [17, 18]. For example, the number of false positives committed by a subject may be subtracted out of their recognition score in order to measure of recognition rates not inflated by guesses. At a group level, this subtraction method can be considered a special case of ANCOVA, with a "beta weight" of 1 applied to the covariate (the covariate being the number of false positives) instead of estimating the beta weight using the least-squares criterion of ANCOVA. Therefore, from this perspective, a connection already exists between this subtraction methodology, ANCOVA and multinomial modeling, in that all three can be used to provide a measure of recognition rate not inflated by guesses. Applied to externalization and internalization biases instead of recognition indices, ANCOVA (or unweighted difference scores) can be considered a method for producing "corrected bias" indices, whereby externalization and internalization biases can be corrected for guessing rates to the extent that guessing rates are reflected by particular types of false positives (as is assumed in multinomial modeling).

At first glance, it may appear that multinomial modeling has an advantage over ANCOVA, because it simultaneously provides a range of estimates of cognitive biases (a, e), all with the number of false positives (b) removed. However, the restrictions that apply to multinomial modeling in terms of the number of estimable parameters do not apply to ANCOVA, providing flexibility. Thus, ANCOVA can also be used to provide a range of adjusted estimates of cognitive biases, and these adjustments can be made for a range of specific types of false positives as opposed to false positives in general (the latter being the only method available in validated multinomial models). False-positive rates are readily available for all sources and are simply the number of trials on which a particular source response is given in response to new items. For example, for externalization errors (i.e., the number of times a person responded with an external source for an internally generated item), guesses can be factored out by using the number of trials on which that subject responded with an external source in response to *new* items as a covariate. The same strategy could be used for any error type of interest, and comparison between groups is commonly made within the ANCOVA framework.

For ANCOVA-based group comparisons using the self-, computer-, and experimenter-generated sources depicted in Figs. 1 and 2 and Table 1, for measures of recognition, the total count of false-positive responses (i.e., experimenter, computer, and self-responses to new items) would be used as a covariate, because these guessing rates can be assumed

to contribute to the recognition measures, just as they contributed to the false positives. For measures of externalization, the count of experimenter and computer responses to new items can be used as a covariate. For measures of internalization (the number of words originating from an external source misremembered as internally generated), the count of self-responses to new items can be used as a covariate. Finally, for the measure of confusion of external sources, the sum of the count of experimenter and computer responses to new item can be used as a covariate.

For a full explanation and employment of these ANCOVA methods, see Woodward, Menon, and Whitman [1]. One exception to this rule is source discrimination scores. False-positive errors are typically *negatively* correlated with source discrimination scores; therefore, high source discrimination scores are not the result of a high guessing rate, and a downward adjustment would not be appropriate. The following is a mathematical depiction of how ANCOVA could be used to test for group differences in externalizations, taking into account guessing external sources:

$$Y = X_1 b_{\text{group}} + X_2 b_{\text{cov}} + E, \qquad (1)$$

where Y = a vector of the mean number of externalization errors (averaged over experimenter and computer for each subject); X_1 = a vector of group membership values (e.g., hallucinating and not hallucinating); X_2 = a vector of the mean number of external false positives (averaged over experimenter and computer for each subject); $[b_{\text{group}}, b_{\text{cov}}]' = (X'X)^{-1}X'Y$ where $X = [X_1, X_2]$. A statistical test of $b_{\text{group}} \neq 0$ tests for an association between hallucinations and the externalization bias, with the influence of guessing removed from externalization biases.

Assumptions associated with ANCOVA

Care must be taken to ensure that ANCOVA assumptions are met in order to produce valid conclusions. All the assumptions associated with standard analysis of variance (ANOVA) also apply to ANCOVA (e.g., normality, independence, homogeneity of variance). In addition, ANCOVA also requires a significance test of each covariate and of homogeneity of regression, all of which are readily available with standard ANCOVA procedures, and the methods for this are explained in detail elsewhere (e.g., pp. 516–517 in [19]). The homogeneity of regression assumption is required for appropriate use of ANCOVA and states that the relationship between the covariate and the dependent variable cannot differ between the groups (i.e., the covariate cannot interact with the treatment), because it involves pooling this regression weight over the two groups.

When using ANCOVA with preexisting, nonrandomly assigned groups, as is the case in the study of schizophrenia and/or hallucinations, particular care must be taken. It is widely agreed that using a covariate to control for a preexisting attribute that differs between the groups is inappropriate (e.g., see [20]). Much of the problem can be subsumed under the heading of "specification errors". As an extreme example, if age-in-years was used as a covariate to adjust old and young groups, this would be analogous to comparing a group of old people and young people that are the same age, which is of course impossible and nonsensical. Conceptually, this is like taking cognitively and physically slow people between 20 and 30 who have been grandparents and have experienced the depression and comparing them to a group of regular people between 20 and 30. In the current application of ANCOVA, this type of error is not a major concern because (1) preexisting attributes are not used as a covariate; instead, rates of false-positive are used as a covariate, and (2) the adjustment does not target the grouping variable but targets the dependent variable and is therefore conceptually similar to studying a difference score, as was mentioned above in the context of the commonly used "corrected recognition" index.

Even without specification errors, care must be taken to ensure that the covariate is not significantly correlated with the grouping variable. If it is, the meaning of this grouping variable may not be clear, because it will be residualized with respect to the covariate [20]. If these circumstances arise (i.e., a significant correlation between the grouping variable and the covariate), it would be necessary to use the difference score methodology as an alternative to ANCOVA and to submit the difference scores to a standard ANOVA. Thus, as long as the covariate is not significantly correlated with the grouping variable, and is significantly correlated with the dependent measure, and the above assumptions are met, we recommend this usage ANCOVA for the study of externalization and internalization biases. A comparison of the assumptions and challenges associated with multinomial modeling and ANCOVA is listed in Table 2.

Method comparison on example data

In Table 1, we have listed the response frequencies for 15 hallucinating schizophrenia patients as a function of source and response, on a source monitoring test with 25 each of self-, computer- and experimenter-generated items and 25 new items (resulting in 100 presented items). In Table 3, we have listed the results of three analyses of these data, in comparison with a sample of 35 nonhallucinating schizophrenia patients (taken from [10]). The three types of analyses are ANCOVA, multinomial modeling with one

Table 2 Comparison of assumptions and challenges associated with each analysis method

Multinomial modeling	ANCOVA
Limitation on number of parameters that can be estimated	Group variables and covariate should be uncorrelated
Goodness of fit required at each step of model development	All assumptions associated with regular analysis of variance apply (e.g., normality, independence, homogeneity of variance).
Model must fit for each individual	Significance test for each covariate required
Validation of models required through cognitive experimentation	Homogeneity of regression required
Selection between 1HT and 2HT models required and no clear guidelines present	

a parameter (Fig. 1), and multinomial modeling with separate a parameters (Fig. 2). In order to adjust the scale of the ANCOVA parameters to equate those derived from multinomial modeling, the values listed for the ANCOVA analysis are estimated marginal means divided by the total number of relevant items. For example, the meaning of a_1 in Table 3 can be derived from Figs. 1 and 2, but for the ANCOVA results some explanation is required. For the multinomial model, a_1 can be interpreted in the context of Fig. 2 and is the probability of responding self when the originating source was the computer or the experimenter. For the ANCOVA model, a_1 also represents the probability of responding self when the originating source was the computer or the experimenter but is computed using estimated marginal means. The estimated marginal mean for naming the self-source for an externally generated item is 0.69 for hallucinating patients and 1.07 for nonhallucinating patients. This number is then divided by the 50 items originating from an external source (25 experimenter generated and 25 computer generated), resulting in the probabilities of 0.01 for hallucinating patients and 0.02 for nonhallucinating patients. With respect to the a_2 parameter, for the ANCOVA model, the estimated marginal mean for naming an external source for self-generated items is 4.05 for hallucinating patients and 2.26 for nonhallucinating patients. Divided by the total number of self-generated

items (25), this resulted in the probabilities of 0.16 and 0.09 for hallucinating and nonhallucinating patients, respectively.

Table 3 highlights some important differences between the results from each model. For example, in the single a parameter model, $a_1 = a_2$ and $e_1 = e_2$, due to the nature of the model (please see Fig. 1 where $a_1 = a_2 = a$ and $e_1 = e_2 = e$). This is precisely the reason why this model cannot be used for the study of hallucinations, where it is necessary to estimate separate error parameters for items that originated from the external source and those that originated from the internal source; that is to say, parameter estimates of biases must be allowed to vary with the originating source. No such restriction need be imposed on the regression model or the multinomial model with separate a parameters (Fig. 2). However, the multinomial model with separate a parameters has not been validated by cognitive experimentation (see assumptions). Note that the probabilities appear much higher for the multinomial models relative to the ANCOVA model in some cases, because the ANCOVA provides the *overall* probability of a given event, but the multinomial model is meant to convey the probability that cognitive processing events in question will take place. The pattern of results indicate that the traditional multinomial model did not detect an externalization bias in the

Table 3 Results' comparison for three methods as a function of hallucinations

Group	D_1	D_2	D_3	d_1	d_2	d_3	a_1	a_2	e_1	e_2	b	g
ANCOVA (estimated marginal means/total relevant items)												
Hallucinating	0.61	0.61	0.86	0.46	0.33	0.54*	0.01	0.16*	0.10	0.12*	0.16	0.08
Nonhallucinating	0.63	0.62	0.86	0.44	0.39	0.68	0.02	0.09	0.08	0.05	0.18	0.09
Multinomial modeling (one a parameter—Fig. 1)												
Hallucinating	0.56	0.53	0.85	0.65	0.00	0.60	0.95	0.95	0.40	0.40	0.18	0.94
Nonhallucinating	0.53	0.53	0.85	0.68	0.09	0.79	0.90	0.90	0.28	0.28	0.19	0.88
Multinomial modeling (separate a parameters—Fig. 2)												
Hallucinating	0.56	0.53	0.85	0.62	0.19	0.38*	0.06	0.61*	0.53	0.35	0.17	0.95
Nonhallucinating	0.53	0.53	0.85	0.59	0.42	0.56	0.13	0.43	0.50	0.27	0.19	0.89

Please see the Fig. 1 and 2 captions for parameter definitions. In the one a parameter model, $a_1 = a_2$ and $e_1 = e_2$, due to the nature of the model (please see Fig. 1). No such restriction need be imposed on the ANCOVA model or the multinomial model with separate a parameters.
* $P < 0.05$ (one tailed) for difference between hallucinating and nonhallucinating patients

hallucinations group, which is expected given the single a parameter as we have discussed in detail above. Both the ANCOVA and the multinomial model with separate a parameters detected a reduction in source discrimination for the internal source for the hallucinating group (d_3) and an associated increase in assignments of internally generated events to an external source (externalizations; a_2). The ANCOVA results also suggested that the externalizations were more frequently assigned to the experimenter than those to the computer for the hallucinating group (e_2), which may be an important result for theoretical accounts suggesting that the cognitive operations underlying hallucinations overlap with those producing errors in internal/external source monitoring.

Summary

We have put forward a number of considerations that must be addressed when attempting to develop new multinomial models and suggest that these often render multinomial modeling impractical. We instead recommend ANCOVA (or difference score methodology if certain ANCOVA assumptions cannot be met), as an appropriate method for removing the influence of external-source guessing rates (external-source false positives) from externalization bias rates. Unlike multinomial models, ANCOVA models are not restricted in terms of the number of estimable parameters, model fit is not a prerequisite for carrying out the analysis, and validation through cognitive experimentation is not required. Note however that care must be taken to ensure that the assumptions of ANCOVA are met, and if they are not, difference score methodology combined with ANOVA may be more appropriate. In addition to source monitoring, the ANCOVA model may also be applied to other cognitive paradigms requiring separation of old–new and source information (e.g. "directed forgetting" paradigms [21]).

References

1. Woodward TS, Menon M, Whitman JC (2007) Source monitoring biases associated with hallucinations in schizophrenia. Cogn Neuropsychiatry 12(6):477–494

2. Ditman T, Kuperberg GR (2005) A source-monitoring account of auditory verbal hallucinations in patients with schizophrenia. Harv Rev Psychiatry 13(5):280–299

3. Larøi F, Woodward TS (2007) Hallucinations from a cognitive perspective. Harv Rev Psychiatry 15(3):109–117

4. Bentall RP, Baker GA, Havers S (1991) Reality monitoring and psychotic hallucinations. Br J Clin Psychol 30:213–222

5. Brébion G, Amador X, David A, Malaspina D, Sharif Z, Gorman JM (2000) Positive symptomatology and source-monitoring failure in schizophrenia-an analysis of symptom-specific effects. Psychiatry Res 95:119–131

6. Larøi F, Van Der Linden M, Marczewski P (2004) The effects of emotional salience, cognitive effort and meta-cognitive beliefs on a reality monitoring task in hallucination-prone subjects. Br J Clin Psychol 43:221–233

7. Batchelder WH, Riefer DM (1990) Multinomial processing models of source monitoring. Psychol Rev 97(4):548–564

8. Batchelder WH, Riefer DM (1999) Theoretical and empirical review of multinomial process tree modeling. Psychon Bull Rev 6(1):57–86

9. Keefe R, Arnold M, Bayen U, McEvoy J, Wilson W (2002) Source-monitoring deficits for self-generated stimuli in schizophrenia: Multinomial modeling of data from three sources. Schizophr Res 57(1):51–67

10. Woodward TS, Menon M, Hu X, Keefe RSE (2006) Optimization of a multinomial model for investigations of hallucinations and delusions with source monitoring. Schizophr Res 85:106–112

11. Bayen UJ, Murnane K, Erdfelder E (1996) Source discrimination, item detection, and multinomial models of source monitoring. J Exp Psychol Learn Mem Cogn 22(1):197–215

12. Riefer DM, Batchelder W (1988) Multinomial modeling and the measurement of cognitive processes. Psychol Rev 95(3):318–339

13. Menon M, Woodward TS (2007) A comparison of one-high-threshold and two-high-threshold multinomial models of source monitoring. J Modern Appl Statist Methods 6(1):279–290

14. Bayen UJ, Nakamura GV, Dupuis SE, Yang CL (2000) The use of schematic knowledge about sources in source monitoring. Mem Cogn 28(3):480–500

15. Smith RE, Bayen UJ (2004) A multinomial model of event-based prospective memory. J Exp Psychol Learn Mem Cogn 30(4):756–777

16. Riefer DM, Batchelder WH (1991) Statistical inference for multinomial processing tree models. In: Doignon J, Falmagne J (eds) Mathematical psychology: current developments. Springer, New York, pp 314–335

17. Nelson NW, Boone K, Dueck A, Wagener L, Lu P, Grills C (2003) Relationships between eight measures of suspect effort. Clin Neuropsychol 17(2):263–272

18. Jacoby LL, Shimizu Y, Daniels KA, Rhodes MG (2005) Modes of cognitive control in recognition and source memory: depth of retrieval. Psychon Bull Rev 12(5):852–857

19. Pedhazur EJ (1982) Multiple regression in behavioral research: explanation and prediction, 2nd edn. Holt, Rinehart and Winston, Toronto

20. Miller GA, Chapman JP (2001) Misunderstanding analysis of covariance. J Abnorm Psychol 110(1):40–48

21. Muller U, Ullsperger M, Hammerstein E, Sachweh S, Becker T (2005) Directed forgetting in schizophrenia prefrontal memory and inhibition deficits. Eur Arch Psychiatry Clin Neurosci 255(4):251–257