Other Papers

Polish Psychological Bulletin
2014, vol 45(3), 363-371
DOI - 10.2478/ppb-2014-0044

*Marek Nieznański **

The role of reinstating generation operations in recognition memory and reality monitoring

***Abstract**: The role of encoding/retrieval conditions compatibility was investigated in a reality-monitoring task. An experiment was conducted which showed a positive effect of reinstating distinctive encoding operations at test. That is, generation of a low-frequency (LF) word from the same word fragment at study and test significantly enhanced item recognition memory. However, reinstating of relatively more automatic operations of reading or generating a high-frequency (HF) word did not influence recognition performance. Moreover, LF words were better recognized than HF words, but memory for source did not depend on the encoding/retrieval match or on the word-frequency. In comparison with reading, generating an item at study significantly enhanced source memory but generating it at test had no effect. The data were analysed using a multinomial modelling approach which allowed ruling out the influence of a response bias on the measurement of memory ability.*
Key words: *transfer-appropriate processing; recognition memory; generation effect; reality monitoring; source memory; multinomial models*

According to the procedural approach (e.g., Kolers & Roediger, 1984; McNamara & Healy, 2000), the cognitive operations used during the study phase of a memory experiment are encoded together with the to-be-remembered item. Therefore, reinstatement of these operations during the test phase of the experiment should provide additional retrieval cues and enhance access to the memory trace of the target item (cf. Rabinowitz, 1990). Similarly, the transfer-appropriate processing (TAP) framework assumes that performance on a memory task is enhanced by increases in the overlap between the processes carried out during encoding and those carried out during test (Morris, Bransford, & Franks, 1977; cf. Horton & Nash, 1999; Mulligan, 1996). However, as pointed out by Nairne (2002) and Surprenant and Neath (2009), TAP in fact only postulates that memory depends on the relation between processing at study and at test and that this relation does not have to be an exact match or even similarity. In other words, processing during study should be *appropriate* for processing at test, but not necessarily the same or similar. For example, processing is appropriate, when the test is sensitive to the feature information strengthened by encoding operations (cf. deWinstanley & Bjork, 1997; deWinstanley, Bjork, & Bjork, 1996). Reinstating the cognitive operations used at study during a test will be critical to memory performance, provided that these operations are themselves important characteristics of a memory trace. If, however, an item is processed automatically (effortlessly[1]) at study, the operations used to encode it may be not successfully recorded in memory, and reinstatement of these operations at test will not enhance performance (cf. Kolers, 1974, 1975). In the current study, it was assumed that reading leads to a more automatic word processing, while generating a word from a word fragment is a more effortful type of processing (e.g., Glisky & Rabinowitz, 1985; Nieznański, 2011), especially for words of low-frequency of occurrence in language. The object of study here was whether encoding/retrieval compatibility influences both memory for the target item and memory for the item's features. These characteristics indicate the particular origin of information, therefore, this aspect of memory performance will be considered from the perspective of the source-monitoring framework (Johnson,

*** Cardinal Stefan Wyszyński University, Institute of Psychology, ul. Wóycickiego 1/3 bud. 14, 01-938 Warsaw, Poland; mnieznanski@wp.pl**

This project was supported by Grant 2011/01/B/HS6/05853 from the National Science Centre in Poland.

[1] Following Hasher and Zacks (1979) we can assume that there is a continuum of attentional requirements among encoding processes. Automatic encoding demands only minimal attentional resources, by contrast effortful processing requires maximal expenditure of resources.

Hashtroudi, & Lindsay, 1993). In the present experiment, the origin of information (source) will refer to whether an item was read or generated during the study.

The present experiment is an extension of the research started by Glisky and Rabinowitz (1985) and Rabinowitz (1990) and continued more recently by Dewhurst and Knott (2010) (see also Dewhurst & Brandt, 2007; Mulligan & Lozito, 2006). Glisky and Rabinowitz, in a set of experiments, showed that the generation effect (i.e., better memory for items generated than read at study; Slamecka & Graf, 1978) can be made larger when generation operations are repeated at test (cf. Nairne & Widner, 1987). Such a result suggested that the compatibility of operations performed at study and test enhances memory, however, reading at both phases of a memory experiment did not provide any benefit. According to Glisky and Rabinowitz, this was due to the relative lack of specificity of reading which is a skilled, automatic process. They also showed that generation of the same missing letters on each presentation (*generate-generate same* condition) provided a recognition memory advantage over generating the same word but with different letters missing at study and test (*generate-generate different* condition). Moreover, the difference between the *generate-generate different* and *generate-read* conditions was not significant. These results indicated that repeating the general kind of operations is not enough to produce a memory benefit, but rather specific operations have to be repeated. This observation was replicated in Rabinowitz (1990) but was not fully supported by Dewhurst and Knott in their experiment with anagrams solution as a generation task (Expt. 2A). They found that an advantage of the *generate-generate* condition over the *generate-read* condition is still present when study and test anagrams had different solution keys. In another experiment (Expt. 2B), they used two types of generation tasks at study and test (i.e., anagram solving and letter completion) and found that the recognition advantage occurred only when the same type of generation task was reinstated at test.

The results of Glisky and Rabinowitz (1985) and Rabinowitz (1990) would seem to be inconsistent with the findings of Mulligan and Lozito (2006) who found that repetition of generation operations at test reduces recognition accuracy. However, as Dewhurst and Knott (2010, see also Dewhurst & Brandt, 2007) demonstrated, this negative reinstatement effect occurs only in studies manipulating read and generate conditions between groups, whereas in studies manipulating them within groups the effect turns out to be positive. Note that the generation effect is similarly moderated by the experimental design. When generate and read items are intermixed within one list, generate items are probably more distinctive than read items, hence, they receive extra attention during study and this contributes to their memory advantage over read items (cf. Begg, Snider, Foley, & Goddard, 1989; Schmidt & Cherry, 1989; Slamecka & Katsaiti, 1987).

Rabinowitz (1990) also explored the effects of cognitive operations reinstatement on memory for item origin. The process of attribution about the origins of remembered information has been named *source monitoring* by Johnson et al. (1993). A special case of a source-monitoring task in which participants have to discriminate items from an internal source (e.g., self-generated words) from those externally derived (e.g., words seen on a computer screen) has been called a reality-monitoring task. The source-monitoring framework assumes that identification of the origin of information is based on differences in the type and amount of characteristics of memories from various sources (cf. Niedźwieńska, 1998; Nieznański, 2008). For example, representations of externally derived information, on average, may have more perceptual, spatial, or temporal details but fewer operational features than representations of internally generated information, and these differences can be used in reality-monitoring decisions (Johnson & Raye, 1981). Therefore, when a word is self-generated, cognitive operations engaged in this task will help in the attribution of this particular word to an internal source. Several studies indicated that an increase in cognitive effort in generating study items results in more effective discrimination of memories of these items from memories derived effortlessly from perception (e.g., Finke, Johnson, & Shyi,1988; Johnson, Raye, Foley, & Foley, 1981; Nieznański, 2011; Rabinowitz, 1989). In the already mentioned study by Rabinowitz (1990), it was assumed that more difficult generation operations are engaged when participants are required to generate medium-frequency-category instances than high-frequency-category instances, and these more difficult operations produce more distinctive records in memory. Although Rabinowitz generally did not confirm better source memory for generated than read items in his reality-monitoring experiments (which Riefer, Chien, & Reimer, 2007, have accounted for by confounding effects of response bias), he observed that taxonomic frequency affected source memory for generated items—that is, origin was better identified for items generated by difficult operations than for items generated by easy operations.

In contrast to the classical study by Rabinowitz (1990), a more comprehensive method of assessing the reality-monitoring performance was used in the present experiment. If a strong response bias influences participants' performance, traditional measures of source memory may lead to unreliable results (e.g., Riefer, et al., 2007; Vogt & Bröder, 2007). In a reality-monitoring task, in the case of uncertainty, participants tend to attribute an item to the external source rather than to the internal source (the self) (Johnson & Raye, 1981, called this the “it-had-to-be-you” effect). This is because of the participants' belief that a generated item would be undoubtedly recognized, thus, if an item induces some uncertainty it must not had been self-generated (e.g., Meiser, Sattler, & von Hecker, 2007). Therefore, in the current study multinomial modelling was used, which is recommended in literature as a method allowing for separate measurement of different cognitive processes and guessing biases in a source memory task (for more details see e.g., Batchelder & Riefer, 1990; Bayen, Murnane, & Erdfelder, 1996; Bröder & Meiser, 2007; Nieznański, 2007).

In summary, the purpose of the present experiment was fourfold: (1) to investigate the effect of encoding operation repetition at test on item recognition memory and how specific this encoding/retrieval match should be to evoke an effect; (2) to study the consequences of the study-test compatibility for reality monitoring (i.e., memory for source); (3) to study the moderating role of operations' automaticity by contrasting more automatic (reading and generation of HF words) with more effortful (generation of LF words) operations; and finally (4) to control, in all the above observations, the influence of response bias by conducting multinomial modelling analyses.

Method

Participants

Sixty undergraduates participated in the experiment. They received bonus points towards final grades for participation. The participants were recruited from the population of second- and third-semester psychology students of Cardinal Stefan Wyszyński University in Warsaw.

Materials and Procedure

Fifty-two HF nouns were used in one experimental session and 52 LF nouns in another session. In each session, an additional 2 words served as buffers at the beginning and 2 at the end of the word list. Among the 52 nouns, 36 were targets and 16 were distracters, which were close in meaning to the target words. All words were 6 to 7 letters long. In the set of HF words, the frequency of occurrences per half a million in language ranged from 42 to 514 (M = 161.2). In the set of LF words, no frequency exceeded 7 occurrences per half a million. For all words, frequency values were taken from the frequency dictionary by Kurcz, Lewicki, Sambor, Szafran, and Woronczak (1990).

Among the 36 target words, 18 were generated and 18 were read at the study phase of the experiment. Among the 18 generated words, 6 were read at test, 6 were generated at test from the same word-fragments as during the study phase and 6 were generated from different fragments. Among the 18 words read at study, 6 were generated at test and 12 were read. Half of the distracters were read at test and another half were generated. Six versions of lists were prepared and counterbalanced across participants so that each word occurred in each study-test condition equally often. Moreover, half of the participants started from the session with HF words and after finishing it, proceeded to the session with LF words as stimuli, while for the second half of the participants, the order of the sessions was reversed. Therefore, read vs. generate condition was manipulated within-list, while word-frequency (HF vs. LF) was manipulated between-lists (although within-participants).

For generate items, one vowel was missing (e.g., *elem_nt*). In the *generate - generate same* condition the same letter was missing in the word presented at the study phase and at the test phase of the experiment. In the g*enerate - generate different* condition a different letter was missing in the words between the phases of the experiment (e.g., *el_ment*). At study, participants were asked to speak the words aloud and try to remember them as well as the way of study (i.e., whether they generated or read the word). At test, participants were instructed to recognise whether the word was generated, read, or whether it was new. The order of stimulus presentation at study and test was random. The presentation rate at study (4 s for each word) and response recording at test were controlled using E-Prime software.

Multinomial Processing Tree Model

The data were analysed using the multinomial processing approach. A model constructed for the purpose of this experiment was based on a two-high-threshold multinomial processing tree (2HT-MPT) model of source monitoring developed by Bayen, et al. (1996). The present model, as shown in Figure 1, contains 7 trees each prepared for a separate type of source defined by experimental conditions; (a) words read at study and at test, (b) words read at study and generated at test, (c) words generated at study and read at test, (d) words generated at study and generated at test from the same fragment, (e) words generated at study and generated at test from a different fragment, (f) new words generated at test, and (g) new words read at test.

(Figure 1 - see page 366-367)

The branches of the trees represent latent cognitive processes resulting in observed responses (the frequencies of the responses have been shown in Appendix). These processes are described by item detection (D), source memory (d) and decision/guessing bias (a, g, b) parameters. The full version of the constructed model contains 17 parameters but only 14 degrees of freedom in the data, therefore, it is not mathematically identifiable and the number of parameters has to be reduced by imposing certain restrictions on parameters. Bayen, et al. (1996, Fig. 4) showed a nested hierarchy of all identifiable submodels of 2HT-MPT model. All of them assume that the parameter describing the probability of detecting that a distractor is new (D_n) is equal to a parameter reflecting old item detection. Moreover, these submodels assume the equality of guessing parameters a and g or, alternatively, that source memory parameters for different sources (d) are equal. A selection of appropriate restrictions for the current experiment is described in the result section.

The goodness of fit of the model to empirical data was tested with the log-likelihood ratio statistic (G^2), which is distributed asymptotically as a χ^2 distribution. An α level of .05 was used for all statistical tests; at this level, G^2 (1) = 3.84 indicates a critical value. All computations were carried out with the *multiTree* computer program (Moshagen, 2010).

Results

No participant failed to generate any HF word. For LF words, however, they failed to generate 5.5% of items at study, what suggests that generation of a LF word was indeed more difficult than HF word. At test, errors occurred

Figure 1. Multinomial processing tree model constructed for data analysis. The model is based on two-high-threshold model developed by Bayen et al. (1996). Rectangles on the left refer to item types (sources), rectangles on the right refer to response types. The branches of the trees represent the following processes: item detection (*D*), source memory (*d*), and response biases (*a, g, b*). The subscripts under parameters refer to specific study or test conditions, that is, *r* refers to reading, *g* to generating, *gs* to generating from the same fragment, *gd* to generating from a different fragment, and *n* to new items.

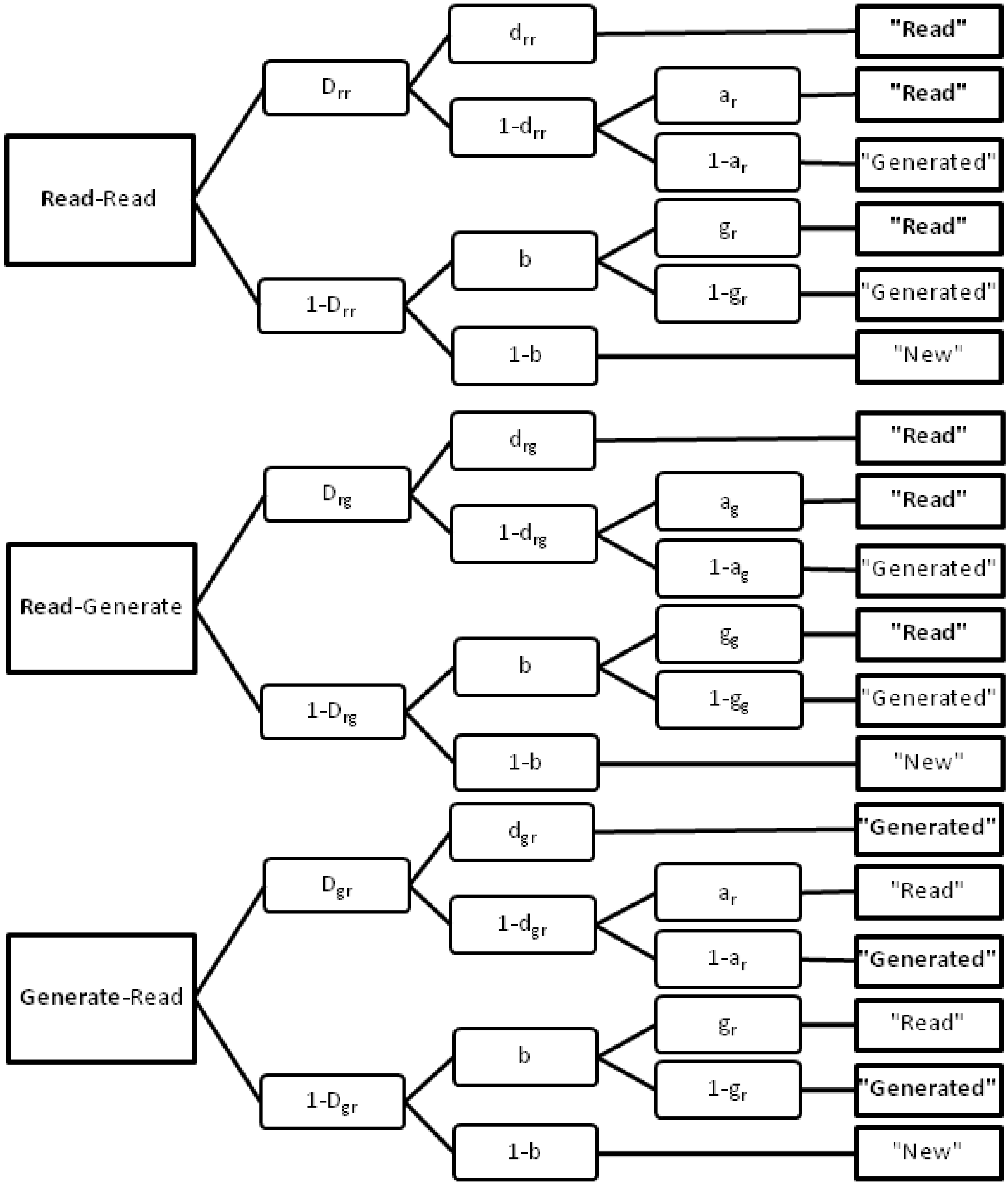
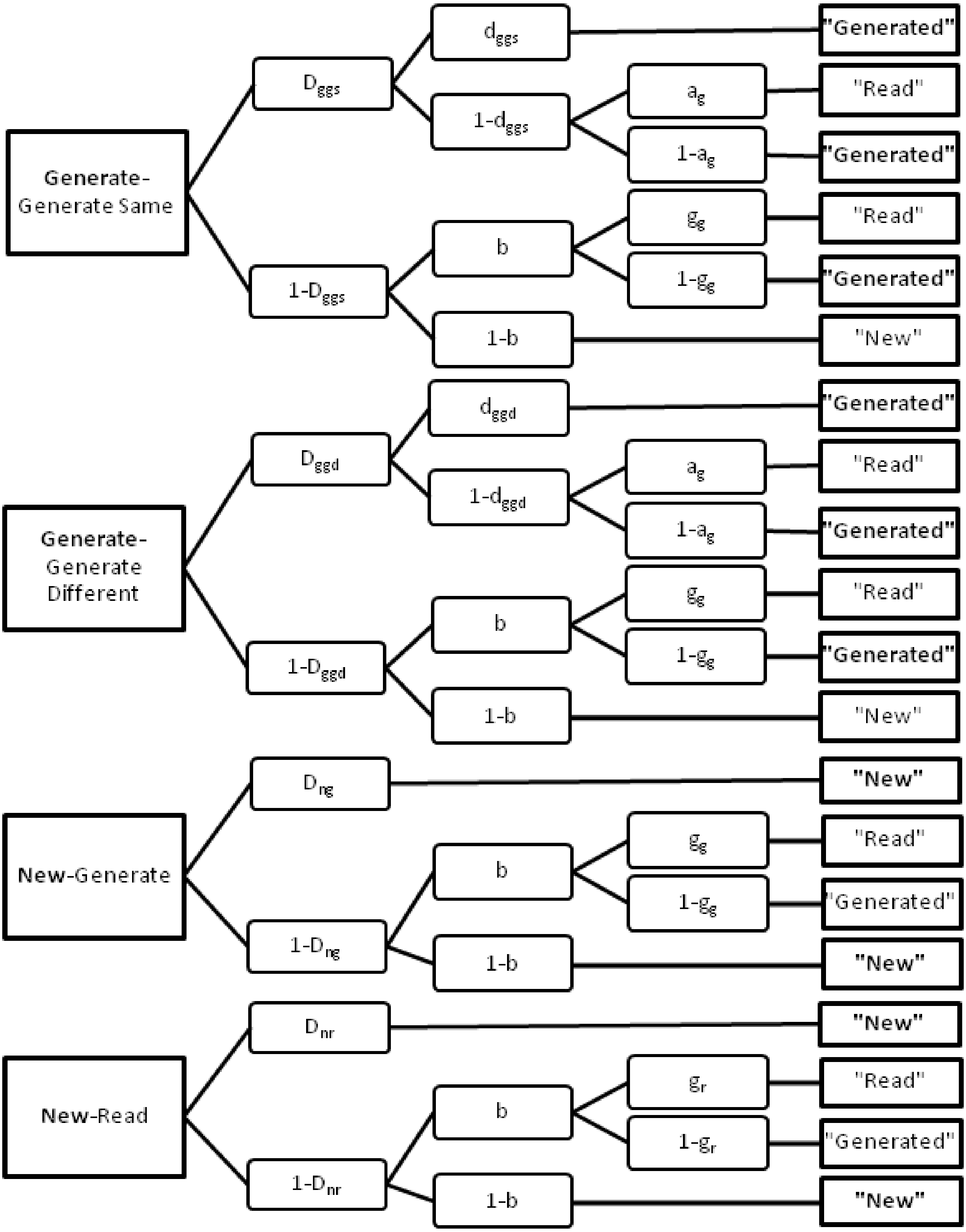


Figure 1. ***part 2***



incidentally and were immediately corrected by the experimenter. Analyses were conducted both on data with and without error trials but they led to the same conclusions concerning item and source memory. The results reported here are based on the analysis of all trials (including error trials).

The same multinomial model was used for the analysis of data gathered for the HF words and LF words and these two models were combined into one general model.

The initial step was to determine the identifiable version of the model which is also acceptable on theoretical grounds. Previous research on the generation effect and reality monitoring suggest that detection parameters and source memory parameters should not be assumed to be equal for read and generate conditions. Similarly, literature on the role of word frequency in recognition memory performance (e.g, MacLeod & Kampe, 1996) also indicates that item detection parameters for LF and HF words should not be assumed equal. Therefore, after considering all possible submodels (Bayen et al., 1996) the following restrictions were imposed; (1) it was assumed that guessing parameters are equal, $a_g = g_g$ and $a_r = g_r$, and do not differ between LF and HF words, and (2) it was assumed that distracters detection parameters are equal to particular detection parameters for old items, that is, $D_{ng} = D_{ggd}$ and $D_{nr} = D_{gr}$. With these assumptions the overall goodness-of-fit of the model was satisfactory, $G^2(4) = 6.88$, $p = .14$. The parameter values of this submodel have been presented on the left side of Table 1. Alternative assumptions concerning distracter detection parameters, $D_{ng} = D_{ggs}$ or $D_{nr} = D_{rr}$, lead to nonsatisfactory fits of such submodels to the data.

In order to explore the role of word frequency in source memory all the parameters reflecting source memory were imposed equal for HF and LF words. A goodness-of-fit test revealed that the overall model fit is still satisfactory after these restrictions, $G^2(9) = 10.03$, $p = .35$. Moreover, when source memory parameters were compared in the pairs for HF vs. LF words for each source separately, the values of $G^2(1)$ never even approached the critical value. This suggests that the word frequency is a variable of no significance for source memory, at least when manipulated between-lists. Therefore, a more simple submodel assuming equality of *d* parameters for HF and LF words was used in further multinomial analyses. The values of this submodel's parameters have been shown on the right side of Table 1.

Item detection

Comparisons between HF words and LF words revealed better item memory (*D*) for LF words than HF words for all types of study-test conditions, the $G^2(1)$ values ranged from 19.42 to 50.19 and were highly significant. Moreover, in nearly all comparisons (made separately for HF and LF words), item memory was significantly lower for items read at study than for items generated at study (the only exception occurred when D_{ggd} was compared with D_{rr} for HR words, in which case the difference was only on a trend level, $G^2(1) = 2.93$, $p = .09$). Next, item memory for words generated from the same fragments (D_{ggs}) was compared with item memory for words generated from different fragments (D_{ggd}). This comparison revealed significantly better memory for words generated from the same fragment, but only in the case of LF words, $G^2(1) =$

Table 1. Parameter estimates and standard errors obtained in multinomial modelling analyses

Parameter	*Model with no restrictions on d parameters*		*Model with equality of d parameters for HF and LF words*	
	HF	*LF*	*HF*	*LF*
$D_{Read\text{-}Read}$	.20 [.08]	.67 [.03]	.20 [.07]	.67 [.03]
$D_{Read\text{-}Generate}$	.26 [.09]	.65 [.04]	.28 [.08]	.65 [.04]
$D_{Generate\text{-}Generate\text{ - }same}$	.59 [.06]	.88 [.02]	.58 [.05]	.88 [.02]
$D_{Generate\text{-}Generate\text{ - }different} = D_{New\text{-}Generate}$	.50 [.03]	.74 [.02]	.51 [.03]	.74 [.02]
$D_{Generate\text{-}Read} = D_{New\text{-}Read}$	.58 [.03]	.76 [.02]	.59 [.02]	.75 [.02]
$d_{Read\text{-}Read}$	.21 [.50]	.39 [.12]	.43 [.10]	
$d_{Read\text{-}Generate}$	.57 [.36]	.38 [.12]	.39 [.12]	
$d_{Generate\text{-}Generate\text{ - }same}$	.70 [.10]	.73 [.04]	.73 [.04]	
$d_{Generate\text{-}Generate\text{ - }different}$	.78 [.09]	.70 [.05]	.73 [.04]	
$d_{Generate\text{-}Read}$	.79 [.07]	.68 [.05]	.70 [.04]	
$a = g_{Generate}$	.65 [.03]		.66 [.03]	
$a = g_{Read}$	.71 [.03]		.70 [.02]	
b	.71 [.02]	.46 [.03]	.70 [.02]	.47 [.03]

13.40, $p < .001$. Item memory for LF words generated at study and read at test (D_{gr}) was significantly worse than for words generated at both phases from the same fragment (D_{ggs}), $G^2(1) = 11.45$, $p < .001$, but it was nearly identical as for words generated from a different fragment at test (D_{ggd}). Item memory for HF words generated at study and read at test (D_{gr}) was very similar to words generated from the same fragments (D_{ggs}) but was significantly better than for words generated from different fragments (D_{ggd}), $G^2(1) = 5.73$, $p < .02$. Comparisons between words read both at study and test (D_{rr}) and words read at study and generated at test (D_{rg}) revealed no significant differences.

Source memory

Significant positive generation effects were found in source memory (d). That is, source was better identified for words generated than for words read at encoding, the $G^2(1)$ values ranged from 5.54 to 9.12, all ps < .05. These effects were found regardless of whether words were generated or read at test. The parameters measuring source memory were nearly identical for words generated from the same fragments (d_{ggs}) and from different fragments (d_{ggd}). There were also no significant differences between source memory parameters for words generated at study and read at test (d_{gr}) and source memory parameters for words generated at both phases (d_{ggs} or d_{ggd}). The same lack of significant differences was observed for words read at study and generated at test (d_{rg}) and words read at both phases (d_{rr}).

Guessing bias

For undifferentiated items, there was a salient tendency to guess that they were read at study. This bias was significantly higher than the neutral value of .50 both for items generated at test, $G^2(1) = 31.59$, $p < .001$, and read at test, $G^2(1) = 43.94$, $p < .001$.

Discussion

The experiment showed an interesting dissociation in the effects of the generation operations reinstatement for HF vs. LF words. In the case of the automatic operation of reading, the encoding/retrieval match had no effect on memory, regardless of the word frequency. The reinstatement of a relatively easy task of generating a HF word by completing one missing letter did not help in item memory but it seems that generation of a HF word by completing a different letter at test disrupted recognition performance. In the case of the more difficult task of generating a LF word, item memory was enhanced when specific operations were reinstated, that is, performance was better for words generated at test from the same fragment than for words read or generated from different fragments. This observation is consistent with the results reported by Glisky and Rabinowitz (1985) and Rabinowitz (1990) but not with the more recent findings of Dewhurst and Knott (2010) who suggested that the reinstatement of specific operations is unnecessary for the reinstatement effect to occur. However, it must be noted that Dewhurst and Knott used a different generation task (anagram solution) than the one applied here and in the Glisky and Rabinowitz research.

There are several studies in literature that have shown a reduction or even lack of the generation effect when unfamiliar stimuli are used in the memory task (McElroy & Slamecka, 1982; Nairne, Pusen, &Widner, 1985). For example, Nairne et al. (1985) observed no generation effect both for nonwords and LF words. To account for these findings they presumed that generation activates an item's representation in a lexical network to a greater degree than does reading. However, in the case of LF words, this activation spreads to relatively few associated entries, hence, there are also relatively few retrieval routes to LF words retained in the memory system. Such an associative linkage hypothesis was challenged by Gardiner, Gregg, and Hampton (1988) research, which showed very similar generation effects for LF and HF words. Also, in the current experiment, a generation effect was found both for LF and HF words.

The current study confirms a well known observation that LF words are better recognized than HF words (e.g, MacLeod & Kampe, 1996). In the case of associative memory, however, several studies reported better performance for HF than LF words (Clark, 1992; Clark & Burchett, 1994; Clark & Shiffrin, 1992). If we consider source memory as a special kind of associative memory, we would predict better source memory for HF than LF words, but such prediction received no support in the results obtained here. The experiment showed that source memory does not depend on word frequency. Such a result may also be interpreted as being inconsistent with the recent experiment on the role of generation difficulty in reality-monitoring performance (Nieznański, 2011, Expt. 1). That experiment showed that memory for an internal source is better for words generated in a difficult than for words generated in an easy condition. Therefore, if we assume that generation of a LF word is more difficult than generation of a HF word, better source memory should be obtained for LF words. However, it must be noted that in that previous experiment a within-list design was used, while in the present experiment LF vs. HF word manipulation was applied between-lists. When difficult and easy trials are mixed within one list, attention may be differently allocated to these two types of trials. When all trials on a list are equally difficult, they are equally distinctive and memory effects between lists are less plausible.

However, read versus generate conditions were manipulated within-lists both in the current experiment and in the previous studies (Nieznański, 2011; Riefer, et al. 2007), and the results concerning the effects of this manipulation are consistent across theses studies. A positive effect of generation on internal source memory was replicated here, that is, words generated at study were correctly identified with greater probability than words read at study. Such a positive generation effect is specific to reality-monitoring tasks – when external sources are used in a source-monitoring task, generation usually disrupts source memory (Mulligan, 2004, 2011; Mulligan, Lozito & Rosner, 2006; Nieznański, 2012, 2013).

In the present experiment, the analysis of a response bias revealed a strong "it-had-to-be-you" effect, that is, participants tended to give a response "read" for unidentified items. It seems that participants aptly believe that generating results in better item memory (cf. deWinstanley & Bjork, 2004), however, this metamemorial belief leads them to a response bias in origin attribution. The multinomial modelling approach allowed ruling out the influence of this response bias from measurement of source memory. The presence of such a response bias constituted a serious problem in the interpretation of the results of Rabinowitz (1990), which were based on traditional indexes of source memory (Batchelder & Riefer, 1990; Riefer et al., 2007).

In conclusion, in the case of more distinctive (less automatic) cognitive operations, their exact repetition at test helps in item recognition. On the basis of the TAP approach, it may be presumed that when a LF word is generated, a distinctive trace is encoded. Hence, performing the same generation task at test provides an appropriate cue facilitating access to the memory trace. However, the encoding/retrieval match does not influence recognition of read words or generated HF words because their encoding operations are relatively effortless and do not provide effective retrieval cues. Implications of these findings for teaching practice are evident and have already been expressed in the concept of 'desirable difficulties' (Bjork, 1994), which suggests that introducing certain difficulties in the training process can enhance the long-term effects of learning. In the case of reality monitoring, both for LF and HF words, reinstatement of encoding operations seem to be insufficient to enhance identification of an item's origin. However, a reality-monitoring task performance is significantly influenced by read vs. generate manipulation but this effect should be measured using methods allowing to control the impact of a response bias.

References

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review, 97*, 548-564.

Bayen U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 197-215.

Begg, I., Snider, A., Foley, F., & Goddard, R. (1989). The generation effect is no artifact: Generating makes words distinctive. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 977-989.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe and A. Shimamura (Eds.). *Metacognition: Knowing about knowing* (pp. 185-205). Cambridge: MIT Press.

Bröder, A., & Meiser, T. (2007). Measuring source memory. *Zeitschrift für Psychologie / Journal of Psychology, 215*, 52-60.

Clark, S. E. (1992). Word frequency effects in associative and item recognition. *Memory & Cognition, 20*, 231-243.

Clark, S. E., & Burchett, R. R. (1994). Word frequency and list composition effects in associative recognition and recall. *Memory & Cognition, 22*, 55-62.

Clark, S. E., & Shiffrin, R. M. (1992). Cuing effects and associative information in recognition memory. *Memory & Cognition, 20*, 580-598.

Dewhurst, S. A., & Brandt, K. R. (2007). Reinstating effortful encoding operations at test enhances episodic remembering. *Quarterly Journal of Experimental Psychology, 60*, 543-550.

Dewhurst, S. A., & Knott, L. M. (2010). Investigating the encoding–retrieval match in recognition memory: Effects of experimental design, specificity, and retention interval. *Memory & Cognition, 38*, 1101-1109.

deWinstanley, P. A., & Bjork, E. L. (1997). Processing instructions and the generation effect: A test of the multifactor transfer-appropriate processing theory. *Memory, 5*, 401-421.

deWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition, 32*, 945-955.

deWinstanley, P. A., Bjork, E. L., & Bjork, R. A. (1996). Generation effects and the lack thereof: The role of transfer-appropriate processing. *Memory, 4,* 31-48.

Horton, K. D., & Nash, B. D. (1999). Perceptual transfer in stem-completion and fragment-completion tests. *Canadian Journal of Experimental Psychology, 53*, 203-218.

Finke, R. A., Johnson, M. K., & Shyi, G. C. -W. (1988). Memory confusions for real and imagined completions of symmetrical visual patterns. *Memory & Cognition, 16*, 133-137.

Gardiner, J. M., Gregg, V. H., & Hampton, J. A. (1988). Word frequency and generation effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 687-693.

Glisky, E. L., & Rabinowitz, J. C. (1985). Enhancing the generation effect through repetition of operations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 193-205.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108,* 356-388.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3-28.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*, 67-85.

Johnson, M. K., Raye, C. L., Foley, H. J., & Foley, M. A. (1981). Cognitive operations and decision bias in reality monitoring. *American Journal of Psychology, 94*, 37-64.

Kolers, P. A. (1974). Two kinds of recognition. *Canadian Journal of Psychology/Revue canadienne de psychologie, 28*, 51-61.

Kolers, P. (1975). Memorial consequences of automatized encoding. *Journal of Experimental Psychology: Human Learning and Memory, 1,* 689-701.

Kolers, P. A., & Roediger III, H. L. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior, 23*, 425-449.

Kurcz, I., Lewicki, A., Sambor, J., Szafran, K., & Woronczak, J. (1990). *Słownik frekwencyjny polszczyzny współczesnej* [Frequency Dictionary of Contemporary Polish]. Kraków: Instytut Języka Polskiego PAN.

MacLeod, C. M., & Kampe, K. E. (1996). Word frequency effects on recall, recognition, and word fragment completion tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 132-142.

McElroy, L. A., & Slamecka, N. J. (1982). Memorial consequences of generating nonwords: Implications for semantic-memory interpretations of the generation effect. *Journal of Verbal Learning and Verbal Behavior, 21,* 249-259.

McNamara, D. S., & Healy, A. F. (1995). A procedural explanation of the generation effect: The use of an operand retrieval strategy for multiplication and addition problems. *Journal of Memory and Language, 34*, 399-416.

Meiser, T., Sattler, C., & von Hecker, U. (2007). Metacognitive inferences in source memory judgments: The role of perceived differences in item recognition. *Quarterly Journal of Experimental Psychology, 60,* 1015-1040.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519-533.

Mulligan, N. W. (1996). The effects of perceptual interference at encoding on implicit memory, explicit memory, and memory for source. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1067-1087.

Mulligan, N. W. (2004). Generation and memory for contextual detail. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 838-855.
Mulligan, N. W. (2011). Generation disrupts memory for intrinsic context but not extrinsic context. *Quarterly Journal of Experimental Psychology, 64*, 1543-1562.
Mulligan, N. W., & Lozito, J. P. (2006). An asymmetry between memory encoding and retrieval. Revelation, generation, and transfer-appropriate processing. *Psychological Science, 17*, 7-11.
Mulligan, N. W., Lozito, J. P., & Rosner, Z. A. (2006). Generation and context memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 836-846.
Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory, 10*, 389-395.
Nairne, J. S., Pusen C., & Widner Jr, R. L. (1985). Representation in the mental lexicon: Implications for theories of the generation effect. *Memory & Cognition, 13*, 183-191.
Niedźwieńska, A. (1998). Identyfikacja źródeł wspomnień [Identification of memories' origins]. *Czasopismo Psychologiczne, 4,* 249-258.
Nieznański, M. (2007). Analiza procesu rozpoznawania źródła informacji w modelach wielomianowych [An analysis of source memory processes according to the multinomial modelling approach]. In: S. Wróbel (Ed.), *Iluzje pamięci* [Memory illusions] (pp. 247-255). Poznań-Kalisz: Wyd. UAM.
Nieznański, M. (2008). Monitorowanie rzeczywistości a subiektywna ocena własności wspomnień wydarzeń i wyobrażeń [Reality monitoring and subjective assessment of remembered events and imaginations' properties]. In: J. Krokos, K. Świętorzecka, & R. Tomanek (Eds.), *W kierunku filozofii klasycznej: inspiracje i kontynuacje* [Towards classical philosophy: inspirations and continuations] (pp. 225-240). Warszawa: Wyd. UKSW.
Nieznański, M. (2011). Generation difficulty and memory for source. *Quarterly Journal of Experimental Psychology, 64,* 1593-1608.
Nieznański, M. (2012). Effects of generation on source memory: A test of the resource tradeoff versus processing hypothesis. *Journal of Cognitive Psychology, 24*, 765-780.
Nieznański, M. (2013). Effects of resource demanding processing on context memory for context-related versus context-unrelated items. *Journal of Cognitive Psychology, 25,* 745-758.
Surprenant, A. M., & Neath, I. (2009). *Principles of memory*. New York: Psychology Press.
Rabinowitz, J. C. (1989). Judgments of origin and generation effects: Comparison between young and elderly adults. *Psychology and Aging, 4,* 259-268.
Rabinowitz, J. C. (1990). Effects of repetition of mental operations on memory for occurrence and origin. *Memory & Cognition, 18*, 72-82.
Riefer, D. M., Chien, Y., & Reimer, J. F. (2007). Positive and negative generation effects in source monitoring. *Quarterly Journal of Experimental Psychology, 60,* 1389-1405.
Schmidt, S. R., & Cherry, K. (1989). The negative generation effect: Delineation of a phenomenon. *Memory & Cognition, 17*, 359-369.
Slamecka, N. J., & Graf, P. (1978). The generation effect: delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 592-604.
Slamecka, N. J., & Katsaiti, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory and Language, 26,* 589-607.
Vogt, V., & Bröder, A. (2007). Independent retrieval of source dimensions: An extension of results by Starns and Hicks (2005) and a comment on the ACSIM measure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 443-450.

Appendix

Response frequencies obtained for high-frequency words

	Response		
Source	*"Read"*	*"Generate"*	*"New"*
Read-Read	400	150	170
Read-Generate	203	79	78
Generate-Read	106	206	48
Generate-Generate-the same	109	208	43
Generate-Generate-different	110	201	49
New-Generate	112	60	308
New-Read	100	38	342

Response frequencies obtained for low-frequency words

	Response		
Source	*"Read"*	*"Generate"*	*"New"*
Read-Read	478	116	126
Read-Generate	223	70	67
Generate-Read	89	216	55
Generate-Generate-the same	69	268	23
Generate-Generate-different	81	237	42
New-Generate	43	22	415
New-Read	31	15	434