Slovenian Version of the Deese–Roediger–McDermott Paradigm: A Normative and Exploratory Study

Jaša Černe* and Urban Kordeš

Center for Cognitive Science, Faculty of Education, University of Ljubljana

Abstract: The research with the Deese–Roediger–McDermott (DRM) paradigm has accumulated substantial evidence for the reliable cross-cultural false memory effect. However, only a limited set of languages has had the benefit of receiving their own version of the DRM paradigm, let alone the corresponding normative data. In this study, we used 728 participants (594 women) to create the first Slovenian version of the DRM paradigm, and another 90 participants (66 women) to test its effectiveness in inducing false memories and conduct exploratory analyses. We are the first one to conduct a normative DRM study entirely online and among the few that measured reaction times (RTs) besides commonly employed accuracy measures. Overall, the participants recalled 69% of the list items and 14% of the critical lures, and additionally recognised 79% of the list items and 45% of the critical lures. The participants recalled and recognised list items as fast as critical lures and significantly faster than unrelated words. The study shows that false memories can be effectively explored by using the online format of the DRM paradigm. Future studies should investigate the relationship between RTs and accuracy measures, the relationship between recall and recognition, as well as the effects of the experimental environment on the results of the DRM studies in more detail. We hope that the provided material and its scrutiny will contribute to the progress in false memory research, especially in the Slovenian language environment.

Keywords: false memory, DRM paradigm, reaction time, online study, normative data

Slovenska različica paradigme Deese–Roediger–McDermott: Normativna in eksploratorna študija

Jaša Černe* in Urban Kordeš

Center za kognitivno znanost, Pedagoška fakulteta, Univerza v Ljubljani

Povzetek: Raziskave s paradigmo Deese–Roediger–McDermott (DRM) so zbrale številne dokaze za zanesljiv medkulturni učinek lažnega spomina. Toda svojo različico paradigme DRM in pripadajoče normativne podatke je prejel le omejen nabor jezikov. V tej študiji smo uporabili 728 udeležencev (594 žensk) za izdelavo prve slovenske različice paradigme DRM in dodatnih 90 udeležencev (66 žensk) za preizkus njene učinkovitosti pri induciranju lažnih spominov in izvedbo eksploratornih analiz. Smo prvi, ki smo normativno študijo DRM izvedli v celoti po spletu, in med redkimi, ki so poleg pogosto uporabljenih mer natančnosti merili reakcijske čase (RČ). Gledano v celoti so udeleženci priklicali 69 % besed s seznama in 14 % kritičnih besed ter dodatno prepoznali 79 % besed s seznama in 45 % kritičnih besed. Udeleženci so priklicali in prepoznali besede s seznama z enako hitrostjo kot kritične besede, za nepovezane besede pa so potrebovali znatno več časa. Študija kaže, da je lažne spomine mogoče učinkovito raziskovati s spletno obliko paradigme DRM. Prihodnje študije bi morale podrobneje raziskati razmerje med RČ in merami natančnosti, razmerje med priklicem in prepoznavo ter vpliv eksperimentalnega okolja na rezultate študij DRM. Upamo, da bodo predloženo gradivo in rezultati analiz prispevali k napredku raziskovanja lažnih spominov, zlasti v slovenskem jezikovnem okolju.

Ključne besede: lažni spomin, paradigma DRM, reakcijski čas, spletna raziskava, normativni podatki

---

*Naslov/Address: Jaša Černe, Center for Cognitive Science, Faculty of Education, University of Ljubljana, Kardeljeva ploščad 16, 1000, e-mail: jasa.cerne@pef.uni-lj.si

The phenomenon of false memory, commonly defined as the recollection of an event that did not occur, has attracted considerable attention in the past few decades. It has been demonstrated in patients with neuropsychological disorders and everyday people of different ages, cultures, and educational levels (for a review, see Brainerd & Reyna, 2005; Gallo, 2006, 2010; Johnson & Raye, 1998). This has led researchers to increasingly adopt the constructivist understanding of memory (e.g., Schacter & Addis, 2007) and challenge important social and legal issues, such as cases of recovered memories of childhood abuse (e.g., Loftus, 1993) or the reliability of eyewitness testimony (e.g., Loftus & Ketcham, 1991). To experiment with the phenomenon, researchers have invented several methods for inducing false memories, one of the most frequently used being the Deese–Roediger–McDermott (DRM) paradigm (Deese, 1959; Roediger & McDermott, 1995). In line with some other recent attempts (Anastasi, DeLeon, & Rhodes, 2005; Iacullo & Marucci, 2016), we present here the first Slovenian version of the DRM paradigm together with the normative and exploratory data obtained entirely online.

In the DRM paradigm, participants are presented with lists of words where the words from each list (i.e., list items) are associatively related to one word that is not presented (i.e., critical lure). When participants are asked to recall or recognise these words, they many times report the critical lure as being part of the list (for reviews, see Gallo, 2006, 2010). For example, list items, such as *hard*, *light*, *pillow*, *plush*, *loud*, etc., often elicit a report of the critical lure *soft* even though it was not presented (Stadler et al., 1999). Typically, critical lures are recalled and recognised less frequently than list items, but more frequently than words that are unrelated to list items (i.e., unrelated lures; Gallo, 2006). Two dominant explanations of the DRM false memory effect are activation-monitoring theory (McDermott & Watson, 2001) and fuzzy-trace theory (FTT; Brainerd & Reyna, 2002). Both maintain that two distinct processes, error-inflating and error-reducing processes, work together in true memory retrieval, but against each other in false memory retrieval (Arndt & Gould, 2006; Gallo, 2010; McDermott & Watson, 2001). Activation-monitoring theory claims that false memories arise because bringing to mind list items repeatedly activates a critical lure, which is then mistaken for a list item if not appropriately monitored. Similarly, FTT assumes that false memories arise because perceiving related list items construct such a strong gist trace (general meaning, usually consistent with the corresponding critical lure) that verbatim traces (specific episodic details) are unable to reject it (Brainerd & Reyna, 2002).

Most findings related to the DRM paradigm are based on English-speaking participants (Anastasi, DeLeon, & Rhodes, 2005; Ulatowska & Olszewska, 2013). Recently, however, several researchers have recognised the need to extend the research to non-English languages. Some researchers directly translated the English word lists (Deese, 1959; Roediger & McDermott, 1995; Stadler et al., 1999) into Chinese (e.g., Lee et al., 2008), Dutch (e.g., Zeelenberg & Pecher, 2002), French (e.g., Cabeza & Lennartson, 2005), German (e.g., Anastasi, Rhodes, et al., 2005), Japanese (e.g., Anastasi, Rhodes, et al.,

2005), Portuguese (e.g., Albuquerque, 2005), and Spanish (e.g., Anastasi, Rhodes, et al., 2005). Yet, there is an obvious problem with direct translations, namely that the translated lists might contain associations that do not exist at all or are uncommon in the target culture and language (Marmolejo et al., 2009). Some researchers thus tried to develop entirely new lists by roughly following the principles outlined by Deese (1959) and Roediger and McDermott (1995) for creating the original lists. To the best of our knowledge, such lists have been developed for Brazilian-Portuguese (e.g., Stein et al., 2006), Italian (e.g., Iacullo & Marucci, 2016), Japanese (e.g., Miyaji & Yama, 2002), Polish (e.g., Ulatowska & Olszewska, 2013), Romanian (e.g., Horoiță & Opre, 2020), Spanish (e.g., Anastasi, De Leon, & Rhodes, 2005), Swedish (e.g., Johansson & Stenberg, 2002), and Turkish (e.g., Göz, 2005). However, no one has yet translated the DRM paradigm directly into Slovenian or created the Slovenian DRM lists *de novo*.

While the DRM paradigm is known for its reliability in inducing spontaneous false memories in the laboratory setting (Brainerd & Reyna, 2005; Gallo, 2006), studies show that this ability varies highly from list to list. In Deese's study (1959), for example, the critical lure *butterfly* was never recalled, whereas the critical lure *sleep* was recalled in 44% of cases. Similarly, Stadler et al. (1999) demonstrated that the critical lure *king* was recalled in only 10% of the cases, whereas critical lure *window* was recalled in 65% of the cases. While these studies were limited to English-speaking environments, the same finding was also observed in studies employing the DRM paradigm in other languages (e.g., Anastasi, De Leon, & Rhodes, 2005). High variability across lists has been connected to forward association strength (FAS), backward association strength (BAS), the structure of the lists, instructions, etc. (Brainerd et al., 2008; Gallo & Roediger, 2002). Notwithstanding, this suggests that the researchers employing the DRM paradigm should be informed about the list-to-list effectiveness of inducing false memories in order to conduct well-designed studies.

Most of the time, the research with the DRM paradigm has focused on categorical accuracy-related dependent variables, such as old/new responses in a recognition test, whereas more continuous measures, such as reaction times (RTs), have been rarely employed, especially with recall tests (Gallo, 2006). This is surprising given that RTs have proven to be a useful objective measure of memory strength (Gallo, 2006; Jou, 2008) that can provide new insight into the cognitive processes underlying false memory, such as separating different stages of false recall (Jou et al., 2004), comparing memory processes between younger and older participants (Tun et al., 1998), or exploring the speed-accuracy trade-off (Coane et al., 2007). Despite the lack of research, the RT data appears to be consistent with the relatedness effect on false memory, according to which critical lures—due to their similarity to studied list items—are recalled and recognised faster than unrelated lures (Gallo, 2006). However, the findings are still inconclusive whether—and under what circumstances—differences exist between RTs of remembering list items and critical lures. Some have demonstrated that the RTs were shorter for recognising list items than for recognising critical lures (Jou et al., 2004; Payne et al., 2002), and similar

results were obtained for the recall data (Jou, 2008). Others, however, found no differences in the recognition RTs between list items and critical lures (Thomas & Sommers, 2005; Tun et al., 1998).

In our study, we developed the Slovenian DRM paradigm from the ground up, tested its effectiveness for inducing false memories, and conducted exploratory analyses by considering both accuracy measures as well as RTs. This is one of the first DRM studies that were conducted online and the first DRM study that obtained extensive normative and exploratory data completely online[1], which is in line with the growing practice of internet-based empirical research (Peirce et al., 2019). Although testing theoretical hypotheses was not our objective, we still had some broad expectations. First, according to the relatedness effect, we expected critical lures would be recalled and recognised more frequently than unrelated lures and less frequently than list items (Anastasi, De Leon, & Rhodes, 2005; Iacullo & Marucci, 2016; Johansson & Stenberg, 2002; Stadler et al., 1999; Ulatowska & Olszewska, 2013). Also, we expected critical lures would be recalled and (positively) recognised either slower (Jou et al., 2004) or equally fast (Tun et al., 1998) as list items, and that both critical lures and list items would be recalled and (positively) recognised faster than unrelated lures (Coane et al., 2007).

Method

Outline of the research design

The study consisted of two stages: (1) creating the Slovenian word lists that can induce false memories, and (2) exploring the effects of the word lists in the normative/exploratory study. In the first stage, we roughly followed the procedures presented in prior research (Deese, 1959; Roediger & McDermott, 1995; Russell & Jenkins, 1954). We translated the words from the Kent–Rosanoff association test (Kent & Rosanoff, 1910), used the translated words to carry out the Slovenian version of the Kent–Rosanoff test, and finally created the word lists. In the second stage, we tested the efficiency of the created word lists to induce false memories, as in the study by Stadler et al. (1999) and conducted exploratory analyses.

Participants

For the first stage of the study (i.e., creating word lists), 1177 volunteers responded to the research participation initiative promoted via mailing lists and various social media groups. Of them, 449 were excluded from the analysis due to not being university students (only students were kept as they represented the majority of the initial sample), having a history of neuropsychological disorders, not being Slovenian native speakers, or providing incomplete data. The remaining

728 (594 women) were used to create the Slovenian word lists. All were university students of various disciplines from the University of Ljubljana, the University of Maribor, and the University of Primorska, aged between 19 and 31 (*M* = 21.7 years, *SD* = 2.2 years). The participants were told that they would take part in the process of creating the Slovenian version of the psychological memory test by providing their associations for a set of words.

Additional 111 volunteers unaware of the first stage of the study were recruited to take part in the second stage of the study (i.e., normative/exploratory part) via mailing lists and various social media groups, but 21 of them were excluded from the analysis due to having a history of neuropsychological disorders, not being Slovenian native speakers, or having experienced technical issues during the study. The remaining 90 (66 women), aged between 18 and 33 (*M* = 22.3 years, *SD* = 3.6 years), were mostly university students (67), but also younger employees (13), unemployed (7), and high school students (2). Out of 67 university students, 37 were promised course credit for their participation. The sample size was determined roughly as per previous research (Anastasi, Rhodes, et al., 2005; Iacullo & Marucci, 2016; Johansson & Stenberg, 2002). The participants were told that they would take part in the process of creating a psychological memory test by listening to audio recordings and completing memory tests, but not that they would be tested for false memories because this could have influenced the results of the study (Gallo et al., 2001). They were fully debriefed after the data collection was completed.

Creating the word lists

Three independent translators translated 100 words from the Kent–Rosanoff association test (Kent & Rosanoff, 1910) to arrive at a basic set of critical lures. Differences in translations were discussed and one translation was selected. Ten words were excluded mostly due to ambiguity (e.g., the word *modra* [*blue*] was ambiguous, for it could mean *being wise* or *a blue colour* in Slovenian), and the remaining 90 words were used in further steps.

Next, an anonymised online questionnaire was used to ask the participants about the first association that comes to mind for each of the 90 words. The obtained associations for each word were ranked by the frequency of occurrence. When very similar associations were given to the same word, the most frequent one was selected and its frequency was replaced by the sum of the frequencies of both associations (e.g., words *svetlo* [*light* or *bright*] and *svetloba* [*light*] were both listed as the associates to the critical lure *okno* [*window*], but only *svetloba* was kept in the final selection as it occurred more frequently in the association test). Very vague, uninformative, or general associations (e.g., word *najboljše* [*the best*] as an association to the word *spanje* [sleep]) were also replaced.

Out of 90 words and their associations, 36 words were selected with the least overlap among their 15 most frequent associations. The remaining overlap was further taken care of by substituting an association that received fewer responses for the most frequent unused association to the same word (e.g., the word *bela* [*white*] was present among the most

---

[1] To the best of our knowledge, only one study has so far employed the online DRM paradigm (Murre et al., 2013), but their goal was not to provide normative data, they did not measure RTs, and they reported only detailed results pertaining to the recognition test.

Figure 1

Data preparation, participant allocation, and study protocol for one participant from Set 1 and one participant from Set 3 (the protocol was equivalent for Set 2)

Data preparation and participant allocation.   Study protocol for one participant from Set$_1$ and one from Set$_3$.
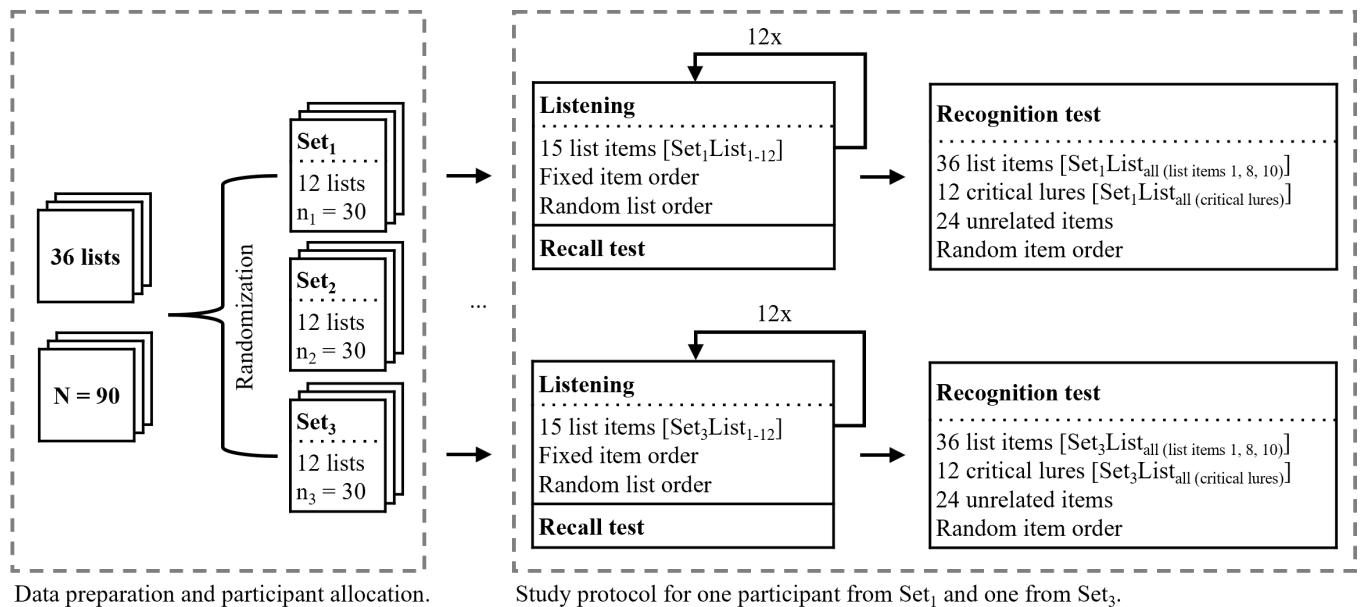
frequent associations to the words *ovca* [*sheep*] and *luna* [*moon*], but the frequency of the second association was lower, so it was substituted for the less frequent association *okrogla* [*round*] to the same word). This led to the development of 36 word lists, with each of them comprising the 15 strongest associations to the list's critical lure (see Appendix A for the complete set of word lists).

Materials

The 36 word lists were audio-recorded and edited so that each recording started with a short beep, followed by 15 words uttered in the descending order of the associative strength[2] with a two-second pause in between, and ended with another short beep. This way, the presentation of the words (e.g., volume, intonation, pace, tone of voice) was identical for all participants. The words were spoken by a male speaker whose native language was Slovenian.

An open-source tool for designing behavioural experiments, PsychoPy v3.0 (Peirce et al., 2019), was used to develop and run an online version of the DRM paradigm. The participants accessed the application via Pavlovia (https://pavlovia.org/), an online repository for running PsychoPy experiments, using their computer, keyboard, and mouse to finish the experiment.

Procedure

Figure 1 summarises the process of data preparation, participant allocation, and study protocol. After agreeing to

take part in the study, the participants received an email with the link to the online socio-demographic questionnaire, a custom-made link to the online experiment, and some general instructions/recommendations on how to proceed. These included: (1) finding a time and place that would minimise potential interruptions; (2) setting up the computer; and (3) testing the speakers or the headphones. The use of either Google Chrome or Mozilla Firefox with no or few tabs open was recommended. The participants were asked to conduct the experiment in one piece, earnestly and with commitment. To prevent tiredness (because of the length of the experiment) from interfering with the results, the 36 lists were divided randomly into three sets of 12 and each participant was pseudo-randomly assigned either to the first, the second or the third set[3]. Since the researcher was not physically present when the participants conducted the online experiment, special care was taken to design the application in a way that helped the participants clearly and unambiguously understand each step in the process.

For the recall test, the participants were told that they would listen to 12 lists of words, each starting and ending with a beep and that after each list, they would have two minutes to type as many words as they could remember. As per previous normative studies (e.g., Stadler et al., 1999), they were told to type only those words for which they were certain were presented in the previous list. The participants were instructed to press the *Return* key if they wished to type another word, or the *Right* key if they could not remember any more words. Each recalled word (i.e., submitted by the *Return* key) appeared in a vertical list on the left side of the screen

---

[2] By associative strength, we mean forward associative strength (FAS), which stands for the probability with which the critical lure elicited the list item on the association test.

[3] Pseudo-randomization allowed us to maintain roughly the same number of participants for each group and to equally distribute the participants, who were promised course credit, into different groups.

in uppercase letters (e.g., "1. MOŽGANI" if the first recalled word was *možgani* [*brain*]). The list with recalled words was shown until the *Right* key was pressed. To approximate the conditions of the usual physical format of the DRM paradigm, the participants were additionally told that they could delete the word if they made a mistake by clicking on the red button next to that word. For each participant, 12 lists of words were presented in random order.

After finishing the recall of the last list from the set, the instructions for the recognition test were presented, followed by the test itself. The recognition test consisted of 72 words: 36 words (i.e., list items) taken from positions 1, 8 and 10 from each of the 12 lists from the set used, 12 critical lures of all 12 lists from the set used, and 24 words taken from the set of words that were excluded in the process of creating the lists (see section *Creating word lists*) and were unrelated to the studied lists (see Appendix B for the complete set of unrelated words used in the recognition test). The words were presented at the centre of the screen in white uppercase letters against a grey background. A blank screen, lasting for one second, was shown between each consecutive word. The participants were instructed to press the *D* key if they were certain that the word was presented in one of the lists, or the *N* key if the word was not presented or if they were not sure whether it was presented. The order of the words was randomised for each participant.

RTs were measured for each recalled and recognised word. Apart from the two minutes timeframe for recalling the words from the individual list, the participants were not told to recall as fast as possible or that RTs are being measured so that no additional (time) pressure was introduced. For the recall test, the RT of the individual word was defined as the time that elapsed from the confirmation of the previous word by pressing the *Return* key (or, for the initial word, when the recall test started) to the confirmation of the current word and so on until two minutes have passed or until the *Right* key was pressed, indicating the end of the recall. For the recognition test, the RT of the individual word was defined as the time that elapsed from the word onset to the output of a response, i.e., pressing the *D* or the *N* key.

Statistical analysis

We used Excel and Python for data preparation and list creation, and the open-source software environment R (R Core Team, 2022) for the statistical analysis.

The exclusion of some participants (see section *Participants*) left us with data from 30 participants for each set of lists. We additionally took care of some fortuitously made errors in the material (e.g., one list had an incorrect number of words on it) and problematic responses in the recall and recognition tests (e.g., participants made several typos when recalling words; see Appendix C for more details). Since we distributed the word lists and the participants to the three sets of lists randomly, we assumed that there should not be any systematic differences in terms of the results of different groups. To improve the statistical power, we therefore combined and jointly analysed the accuracy and RT data from all three sets of lists (see Appendix D and E for descriptive statistics of each set of lists individually). As some

participants did not recall or recognise all item types (i.e., list items, critical lures, and unrelated lures), the final RT dataset was unbalanced, comprising 34 incomplete participant cells for recall, and 40 incomplete participant cells for recognition.

For the analysis of accuracy data, we calculated recall and recognition rates per participant and word list for each item type. To get the participant's list item recall/recognition rates, we divided the number of all correctly recalled/recognised list items of the participant by the number of all list items of all lists that the participant listened to (recall rates) or by the number of all list items that were included in the recognition test (recognition rates). To get the participant's recall/ recognition rates of the critical lures, we divided the number of all recalled/recognised critical lures of the participant by the number of lists that the participant listened to. We used a similar procedure for the calculation of unrelated lure rates and rates per list. For the RT analysis, we calculated average recall and recognition RTs for each item type and response to the recognition test (i.e., yes, no). The focus of the analysis was on testing the differences and the relationships (between item types, responses, and recall/recognition tests) within accuracy data and RTs separately, as well as the relationships between them.

Since the data violated the assumptions of common parametric procedures (e.g., the accuracy and RT distributions differed from normal, some RTs were unusually long, etc.), we decided to use robust statistical methods from the R packages MANOVA.RM (Friedrich et al., 2019) and WRS2 (Mair & Wilcox, 2020), developed to maintain the power in the presence of outliers and non-normal distributions[4]. We used a robust repeated measures ANOVA (RM function from MANOVA.RM) to compare recognition RTs across all combinations of item type and response variable levels (list item-yes, list item-no, critical lure-yes, critical lure-

---

[4] Before using robust procedures, we considered trimming the RTs by using an absolute lower threshold of 100ms and an upper threshold of three standard deviations above the mean separately for each item type (on the recall and recognition test) and response (on the recognition test). Yet, as pointed out by one of the reviewers, there are many problems with trimming the data according to rules of thumb (e.g., Field & Wilcox, 2017). Also, the unusually long RTs in the present study (e.g., the longest recall RTs were 138914 ms, 122552 ms, 118913 ms, etc., and the longest recognition RTs were 99533 ms, 46303 ms, 35684 ms, etc.) could have been the consequence of the online format of the study. Namely, the participants could have experienced internet connection problems, got distracted easier because of the physical absence of the researcher, or taken a little more time for deliberation since they were not instructed to recall and recognise the words as fast as possible. Accordingly, visual inspection suggested that the possible differences between the RTs of list items, critical lures, and unrelated lures exist in the tails of the corresponding distributions. Thus, trimming could indeed skew the data and robust procedures were probably a better choice. We thank the anonymous reviewer for this suggestion.

[5] Since the RM function does not work with unbalanced designs, we balanced the recognition RT dataset by using the complete cases approach just for this procedure, removing 40 (i.e., 44%) participant cells. Each condition (list item-yes, list item-no, critical lure-yes, critical lure-no, unrelated lure-yes, unrelated lure-no) comprised 50 participant cells.

Table 1

Descriptive Statistics for All the Measures Used in the Inferential Statistics of Accuracy and RT Data

		Rates		Reaction times		
		Recall	Recognition	Recall	Recognition (positive)	Recognition (negative)
List items	*N*	90	90	90	90	90
	M	.69	.79	3982.5	1565.4	2239.6
	SD	.11	.11	1388.7	732.0	1628.6
	Range	.58	.49	8230.5	4752.8	12672.2
	Skewness	0.00	−0.50	1.67	2.69	4.15
	Kurtosis	0.34	−0.27	3.98	9.59	24.29
Critical lures	*N*	90	90	68	87	88
	M	.14	.45	5848.3	1638.7	3186.2
	SD	.13	.25	14520.7	1127.9	3872.1
	Range	.55	1.00	117993.3	8726.8	33409.3
	Skewness	0.99	0.08	6.96	4.18	6.20
	Kurtosis	0.40	−0.72	51.09	24.05	45.04
Unrelated lures	*N*		90	69	54	90
	M		.08	9267.8	2528.3	1552.5
	SD		.08	12202.2	2038.7	645.9
	Range		.33	70368.0	11368.0	3652.8
	Skewness		1.02	3.67	2.66	2.10
	Kurtosis		0.71	15.36	8.41	5.25

Note. Because the maximum number of recalled unrelated lures was not known in advance (as it was with recall/recognition list items and critical lures), we could not calculate comparable rates. *N* = the number of participants with at least one response in a given category.

no, unrelated lure-yes, unrelated lure-no)[5]. We report a resampling version of Wald-type statistic (WTS) based on a permutation method. For multiple within-subject comparisons within one variable, we used a bootstrap-based repeated measures ANOVA for trimmed means (*rmanovab* function from WRS2) and the corresponding *post hoc* tests (*pairdepb* function from WRS2). Because the *pairdepb* function does not work with unbalanced designs, we sometimes used its non-bootstrap alternative (*rmmcp* function from WRS2). For *post hoc* tests, we report a difference between trimmed means ($\hat{\psi}$), an associated bootstrap confidence interval, and the bootstrapped effect size (δ_t; Algina et al., 2005), where 0.2, 0.5, and 0.8 correspond to small, medium, and large effects. For comparing two dependent means, we used Yuen's modified *t*-test for trimmed means (*yuend* function from WRS2) and bootstrapped effect size (Algina et al., 2005). Finally, for correlational analyses, we used R package *boot* to implement Pearson's correlation with bootstrapping. We report bias-corrected and accelerated (BCa) bootstrap confidence intervals.

As per recommendations (Field & Wilcox, 2017), we used 20% trimming for all trimmed means procedures and 2000 samples for all bootstrap-based procedures. The level for assessing the statistical significance was .05.

Results

Table 1 shows descriptive statistics for all datasets used in the inferential procedures.

Results of accuracy data

Recall accuracy

All participants recalled at least one list item, 67 out of 90 participants recalled at least one critical lure, and 69 out of 90 participants recalled at least one unrelated lure. Of all recalled words, 97.16% were list items, 1.32% were critical lures, and 1.51% were unrelated lures.

Table 2 shows recall rates of critical lures and list items for each list. The recall rates of list items varied from 80% (*sadje* [*fruit*]) to 54% (*kratko* [*short*]), whereas the recall rates of critical lures were lower and varied from 47% (*zdravnik* [*doctor*]) to 0% (e.g., *štedilnik* [*stove*]). Robust paired Yuen's test showed the participants were more likely to recall list items than critical lures, $M_{diff} = 0.57$, 95% CI [0.53, 0.62], $Y_t(53) = 27.04$, $p < .001$, $\delta_t = 3.48$.

Table 2
Recall Rates of Critical Lures and List Items for the 36 Slovenian Word Lists

| | Top 18 lists | | | | | Bottom 18 lists | | | |
| | List items | | Critical lures | | | List items | | Critical lures | |
List	M	SD	M	SD	List	M	SD	M	SD
Zdravnik (doctor)	.68	.16	.47	.51	Dekle (girl)*	.70	.13	.10	.31
Gladko (smooth)	.63	.17	.37	.49	Lev (lion)	.68	.17	.10	.31
Okno (window)	.66	.14	.33	.48	Mehko (soft)	.70	.13	.10	.31
Trg (square)	.70	.12	.33	.48	Otrok (child)	.72	.16	.10	.31
Jeza (anger)	.65	.14	.30	.47	Pajek (spider)	.68	.19	.10	.31
Strah (afraid)	.64	.15	.30	.47	Glasba (music)	.73	.17	.07	.25
Spanje (sleep)	.58	.12	.27	.45	Gora (mountain)	.72	.12	.07	.25
Želja (wish)	.62	.16	.27	.45	Ovca (sheep)	.77	.13	.07	.25
Sir (cheese)	.78	.14	.20	.41	Moški (man)	.73	.12	.03	.18
Glava (head)	.70	.17	.17	.38	Sadje (fruit)	.80	.12	.03	.18
Igla (needle)	.65	.17	.17	.38	Sladko (sweet)	.69	.17	.03	.18
Metulj (butterfly)	.63	.19	.17	.38	Sol (salt)	.68	.11	.03	.18
Rdeče (red)	.65	.15	.17	.38	Visoko (high)	.77	.14	.03	.18
Kratko (short)	.54	.16	.13	.35	Bolezen (sickness)	.70	.17	.00	.00
Luna (moon)	.69	.17	.13	.35	Cvet (blossom)	.69	.17	.00	.00
Reka (river)	.76	.12	.13	.35	Dolgo (long)	.69	.15	.00	.00
Ulica (street)	.71	.16	.13	.35	Globoko (deep)	.69	.13	.00	.00
Zelje (cabbage)	.76	.13	.13	.35	Štedilnik (stove)	.62	.19	.00	.00
Total (all lists)	.69	.11	.14	.13					

Note. The lists are ordered in descending order of the mean critical lure rates. By accident, the list with the critical lure *dekle* contained 16 instead of 15 words, hence its results are excluded from the total *M* and *SD* calculation.

Recognition accuracy

All participants recognised at least one list item, 87 out of 90 participants recognised at least one critical lure, and 53 out of 90 participants recognised at least one unrelated lure. Of all recognised words, 81.64% were list items, 15.64% were critical lures, and 2.72% were unrelated lures.
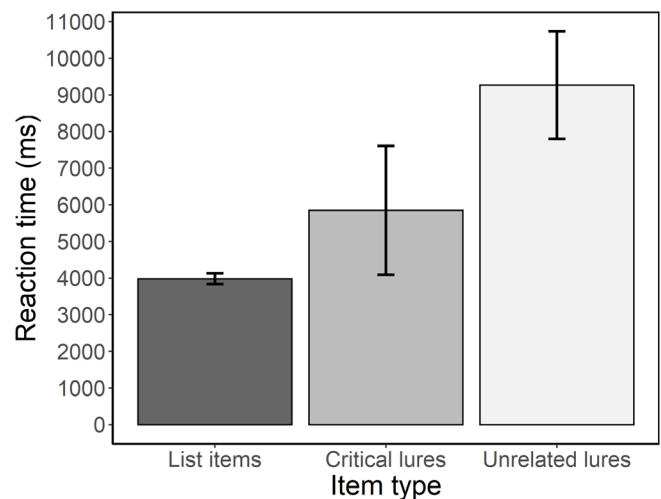
Table 3 shows recognition rates of critical lures and list items for each list. We can see that the rates of list items varied from 99% (*sadje* [*fruit*]) to 68% (e.g., *kratko* [*short*]), whereas the rates of critical items were again lower and varied from 77% (*trg* [*square*]) to 13% (*lev* [*lion*]). Robust paired Yuen's test showed the participants were more likely to recognise list items than critical lures, $M_{diff} = 0.35$, 95% CI [0.29, 0.41], $Y_t(53) = 11.04$, $p < .001$, $\delta_t = 1.09$, and unrelated lures, $M_{diff} = 0.74$, CI [0.70, 0.77], $Y_t(53) = 40.48$, $p < .001$, $\delta_t = 4.71$. They were also more likely to recognise critical lures than unrelated lures, $M_{diff} = 0.38$, CI [0.33, 0.44], $Y_t(53) = 14.75$, $p < .001$, $\delta_t = 1.57$.

Comparison of recall and recognition accuracy

As indicated by the robust paired Yuen's tests, the comparison of recall and recognition rates showed that the

Figure 2
Differences in recall RTs of list items, critical lures, and unrelated lures

Note. Error bars depict the standard errors of the mean.

Table 3

Recognition Rates of Critical Lures and List Items for the 36 Slovenian Word Lists

| | Top 18 lists | | | | | Bottom 18 lists | | | |
| | List items | | Critical lures | | | List items | | Critical lures | |
List	M	SD	M	SD	List	M	SD	M	SD
Trg (square)	.91	.15	.77	.43	Želja (wish)	.70	.27	.43	.50
Okno (window)	.60	.32	.73	.45	Bolezen (sickness)	.82	.24	.40	.50
Mehko (soft)	.80	.22	.70	.47	Gora (mountain)	.77	.19	.40	.50
Spanje (sleep)	.73	.25	.70	.47	Cvet (blossom)	.85	.16	.37	.49
Zdravnik (doctor)	.78	.24	.70	.47	Moški (man)	.73	.25	.37	.49
Gladko (smooth)	.70	.27	.67	.48	Sladko (sweet)	.85	.16	.37	.49
Jeza (anger)	.79	.24	.67	.48	Visoko (high)	.74	.26	.37	.49
Igla (needle)	.80	.19	.63	.49	Sir (cheese)	.87	.19	.33	.48
Strah (afraid)	.78	.27	.63	.49	Luna (moon)	.74	.24	.30	.47
Dekle (girl)*	.91	.17	.60	.50	Pajek (spider)	.87	.22	.27	.45
Ovca (sheep)	.80	.22	.60	.50	Glasba (music)	.72	.29	.23	.43
Rdeče (red)	.88	.14	.57	.50	Kratko (short)	.68	.24	.23	.43
Zelje (cabbage)	.92	.14	.53	.51	Metulj (butterfly)	.72	.28	.23	.43
Reka (river)	.87	.15	.50	.51	Sadje (fruit)	.99	.06	.23	.43
Ulica (street)	.82	.27	.50	.51	Dolgo (long)	.80	.24	.20	.41
Otrok (child)	.84	.19	.47	.51	Sol (salt)	.79	.22	.20	.41
Glava (head)	.74	.26	.43	.50	Štedilnik (stove)	.71	.26	.20	.41
Globoko (deep)	.71	.23	.43	.50	Lev (lion)	.77	.18	.13	.35
Total (all lists)	.79	.11	.45	.25					

Note. The lists are ordered in descending order of the mean critical lure rates. Total *M* and *SD* are calculated based on all sets of lists. By accident, the list with critical lure *dekle* had 16 instead of 15 words on it, hence its results are excluded from the total *M* and *SD* calculation.

Table 4

Mean Rates for Critical Lures and List Items Recalled and Mean Rates for Critical Lures and List Items Recognised for the Top 18 Lists from the Present Study, and from the English (Stadler et al., 1999), Swedish (Johansson & Stenberg, 2002), Spanish (Anastasi, De Leon et al., 2005), Polish (Ulatowska & Olszewska, 2013) and Italian (Iacullo & Marucci, 2016) Studies

| | Recall | | | | Recognition | | | |
| | List items | | Critical lures | | List items | | Critical lures | |
Language	M	SD	M	SD	M	SD	M	SD
Slovenian	.67	.06	.23	.10	.80	.08	.60	.11
Italian	.62	.06	.27	.13	.60	.19	.87	.06
Spanish	.52	.06	.39	.16	.77	.08	.79	.12
Polish	.62	.08	.43	.17	.77	.09	.74	.10
English	.59	.05	.51	.07	.71	.07	.77	.06
Swedish	.53	.05	.65	.14			.88	.06
Total	.60	.06	.41	.13	.73	.10	.78	.09

Note. The data is ordered in descending order of the mean critical lure recall rate for the Slovenian lists. The Polish and Swedish studies did not report Ms and SDs for recalled and recognised list items. However, the authors generously provided all missing data except for the Ms and SDs for the recognised list items in the Swedish study.

recognition rates were significantly higher for list items, M_{diff} = 0.10, 95% CI [0.08, 0.12], $Y_t(53) = 9.74$, $p < .001$, $\delta_t = 1.23$, and for critical lures, $M_{\text{diff}} = 0.33$, CI [0.28, 0.37], $Y_t(53) =$ 14.16, $p < .001$, $\delta_t = 1.23$.

Comparison of accuracy data with similar normative studies

We averaged the list item and critical lure rates for the top 18 lists (upper half) from the present study and the five most similar studies that provided norms for English (Stadler et al., 1999), Swedish (Johansson & Stenberg, 2002), Spanish (Anastasi, De Leon, & Rhodes, 2005), Polish (Ulatowska & Olszewska, 2013) and Italian (Iacullo & Marucci, 2016). Table 4 shows the results. Considering critical lures, the present study showed the lowest recall and recognition rates (23% and 60%, respectively), the Swedish study had the highest recall rate (65%), and both the Swedish and the Italian studies had the highest recognition rate (87%). Considering list items, the Spanish study showed the lowest recall rate (52%), the Italian study had the lowest recognition rate (59%), and the present study had the highest recall and recognition rates (67% and 80%, respectively).

Results of reaction times analysis

Recall reaction times

In terms of recall data, a robust one-way repeated measures ANOVA showed RTs differ across item types, $F_t = 11.85$, $F_{\text{crit}} = 3.75$, $p < .05$. On average, as can be seen in Figure 2, the participants recalled list items slightly faster than critical lures, and almost three seconds faster than unrelated lures. *Post hoc* tests (non-bootstrap version) showed statistically significant differences for list items and unrelated lures, $\hat{\psi} = -2659.47$, 95% CI [−4916.77, −402.17], $p_t < .05$, $\delta_t = 0.41$, and for critical lures and unrelated lures, $\hat{\psi} = -2787.36$, CI [−5273.67, −301.06], $p_t < .05$, $\delta_t = 0.40$, but not for list items and critical lures, $\hat{\psi} = 680.27$, CI [−312.33, 1672.87], $p_t > .05$, $\delta_t = 0.29$. One possible reason for the significant differences
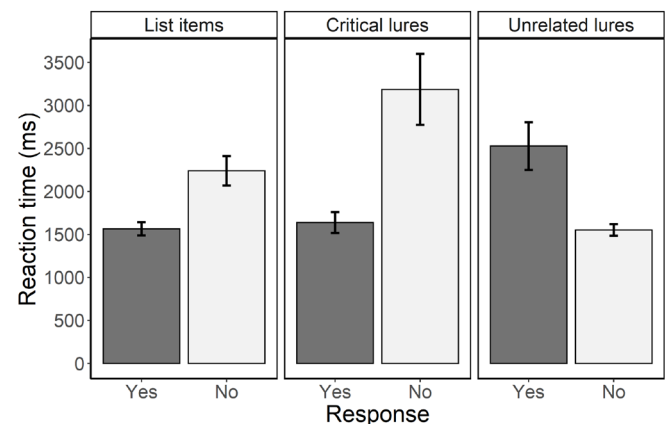
between the recall RTs of different item types is that recalled unrelated lures comprised longer words, thus requiring more time for typing them. However, the comparison of the average length (number of letters) of all the recalled words across different item types did not support such speculation: the length of both recalled list items ($M = 6.0$, $SD = 1.1$) and critical lures ($M = 6.0$, $SD = 1.2$) was even slightly greater than the length of the unrelated lures ($M = 5.6$, $SD = 1.3$).

Recognition reaction times

Figure 3 shows the recognition RTs for positive (pressing the *D* key as a response to the presented word indicating that the word *was* included in one of the lists) and negative (pressing the *N* key as a response to the presented word indicating that the word *was not* included in one of the lists) responses across

Figure 3

Differences in mean RTs for positive and negative responses to the recognition test across list items, critical lures, and unrelated lures

Note. Error bars depict the standard errors of the mean.

Table 5

Results of Robust Pairwise Comparisons (Non-bootstrap Version) Between the Recognition RTs of Different Item Types for Positive and Negative responses

Comparisons	$\hat{\psi}$	95% CI	*p*-value
Positive responses			
List items – Critical lures	70.04	[−81.55, 221.62]	not significant
List items – Unrelated lures	−699.53	[−1201.90, −197.15]	< .05
Critical lures – Unrelated lures	−658.81	[−1108.27, −209.36]	< .05
Negative responses			
List items – Critical lures	−391.34	[−671.08, −111.60]	< .05
List items – Unrelated lures	483.68	[318.83, 648.54]	< .05
Critical lures – Unrelated lures	944.22	[685.18, 1203.25]	< .05

Note. $\hat{\psi}$ = difference between trimmed means (psihat).

different item types. A robust two-way repeated measures ANOVA (complete cases) showed a significant interaction between item types and responses, WTS(2) = 33.24, $p < .001$, indicating that when the participants responded to list items, critical lures, and unrelated lures, the differences in RTs were dependent on whether they responded positively or negatively. More concretely, for list items and critical lures, the RTs were shorter for the positive responses compared to the negative ones, but the opposite was true for unrelated lures.

We conducted *post hoc* tests by comparing (i) different item types within positive and negative responses separately, and (ii) by comparing different responses within list items, critical lures, and unrelated lures separately. For the first comparison, a robust one-way repeated measures ANOVA showed significant differences between different item types for positive, $F_t = 15.60$, $F_{crit} = 4.45$, $p < .05$, and negative responses, $F_t = 45.01$, $F_{crit} = 3.18$, $p < .05$, so we proceeded with further robust pairwise comparisons (non-bootstrap version; Table 5). In terms of positive responses, the RTs of list items did not differ from the RTs of critical lures, however, the RTs of list items and critical lures were significantly shorter compared to unrelated lures. As for negative responses, the results showed that the participants were the slowest at correctly rejecting critical lures, faster at wrongly rejecting list items, and the fastest at correctly rejecting unrelated lures.

As far as the differences between positive and negative responses within individual item types are concerned, a robust paired Yuen's test showed that the RTs for positive responses

Table 6

Results of Correlation Analysis Within Accuracy Data, Within RTs, and Between Accuracy Data and RTs

	Comparisons	r	95% BCa *CI*
Accuracy data	Recall		
	List items vs Critical lures	−.06	[−0.24, 0.15]
	Recognition		
	List items vs Critical lures	−.16	[−0.36, 0.01]
	List items vs Unrelated lures	−.13	[−0.34, 0.07]
	Critical lures vs Unrelated lures	.31*	[0.10, 0.47]
	Recall vs Recognition		
	List items	.65*	[0.49, 0.76]
	List items (word lists)	.56*	[0.28, 0.76]
	Critical lures	.62*	[0.46, 0.73]
	Critical lures (word lists)	.69*	[0.47, 0.81]
RTs	Recall		
	List items vs Critical lures	.40*	[0.07, 0.72]
	List items vs Unrelated lures	.24	[−0.06, 0.48]
	Critical lures vs Unrelated lures	.02	[−0.11, 0.32]
	Recognition		
	List items vs Critical lures	.42*	[0.18, 0.73]
	List items vs Unrelated lures	.08	[−0.14, 0.32]
	Critical lures vs Unrelated lures	.38*	[0.04, 0.73]
	Recall vs Recognition		
	List items	.34*	[0.04, 0.55]
	Critical lures	.35	[−0.06, 0.78]
	Unrelated lures	.12	[−0.17, 0.38]
Accuracy data vs RTs	Recall		
	List items	.18	[−0.03, 0.35]
	Critical lures	−.08	[−0.23, 0.17]
	Recognition		
	List items	−.03	[−0.19, 0.13]
	Critical lures	−.39*	[−0.52, −0.23]
	Unrelated lures	.15	[−0.12, 0.37]

Note. r = Pearson's correlation coefficient; 95% BCa *CI* = 95% bias-corrected and accelerated bootstrap confidence interval; * = significant correlation according to 95% BCa *CI*.

were significantly shorter than the RTs for the negative ones within list items, $M_{\text{diff}} = 473.88$, 95% CI [314.40, 633.37], $Y_t(53) = 5.96$, $p < .001$, $\delta_t = 0.73$, and critical lures, $M_{\text{diff}} = 948.18$, CI [689.02, 1207.33], $Y_t(50) = 7.35$, $p < .001$, $\delta_t = 0.86$, whereas within unrelated lures, the opposite pattern emerged—the RT for positive responses were significantly longer than the RTs for the negative ones, $M_{\text{diff}} = 653.81$, CI [295.19.26, 1012.42], $Y_t(33) = 3.71$, $p < .001$, $\delta_t = 0.41$.

Considering the RTs for correct responses only (compare List items–Yes, Critical lures–No, and Unrelated lures–No in Figure 3), a robust one-way repeated measures ANOVA showed that they varied significantly across item types, $F_t = 75.47$, $F_{\text{crit}} = 3.96$, $p < .05$. Robust pairwise comparisons (non-bootstrap version) showed that the RTs for correct responses to list items (Yes) and unrelated lures (No) were roughly equal, $\hat{\psi} = -40.79$, 95% CI [–123.44, 41.87], whereas the RTs for correct responses to critical lures (No) were significantly longer than the RTs for correct responses to list items, $\hat{\psi} = -1011.27$, CI [–1255.53, –767.01], and to unrelated lures, $\hat{\psi} = 944.18$, CI [685.18, 1203.25].

Results of Correlational Analysis

The results of the correlational analysis are presented in Table 6.

Correlations within accuracy data

In terms of accuracy rates per participant, we could not find any correlation between recalled list items and critical lures, as well as between recognised list items and critical lures, or between recognised list items and unrelated lures. However, we found a mild positive correlation between the rates of recognised critical and unrelated lures. Concerning participants' recall and recognition rates, we found a moderate positive correlation between the recall and recognition rates both for list items and critical lures. In terms of data averaged by word lists, we observed a similar positive correlation between the recall and recognition rates for list items and critical lures.

Correlations within reaction times

Regarding participants' recall RTs across item types, we found a moderate positive correlation between list items and critical lures, but not between list items and unrelated lures or between critical and unrelated lures. In other words, the participants who recalled list items faster also took less time when recalling critical lures.

As with recall RTs, the analysis of recognition RT showed a moderate positive correlation between list items and critical lures, and no correlation between list items and unrelated lures. However, contrary to the recall RTs, there was a mild positive correlation between recognition RTs of critical and unrelated lures, which indicates that the participants who recognised critical lures faster also did so for unrelated lures.

Considering recall RTs and RTs of positive responses to the recognition test, we found a mild positive correlation for list items, but no correlation for critical lures or unrelated lures, indicating that the participants who recalled list items faster also recognised them faster.

Correlations between accuracy data and reaction times

We found no correlation between recall rates and RTs for list items or critical lures. Likewise, we could not find a correlation between recognition rates and RTs for list items or unrelated lures; however, we found a mild negative correlation for critical lures, which indicates that the participants who responded faster to the onset of critical lures in the recognition test were also more likely to respond to critical lures positively.

Discussion

The purpose of the present study was to develop the first Slovenian (online) version of the DRM paradigm, explore its effectiveness to induce false memories and provide exploratory analyses. In the remaining part of the paper, we comment on the obtained results and relate them to prior literature.

We demonstrated a similar, albeit lower false memory effect to that of prior research (Anastasi, De Leon, & Rhodes, 2005; Iacullo & Marucci, 2016; Johansson & Stenberg, 2002; Stadler et al., 1999; Ulatowska & Olszewska, 2013). According to our expectation, the participants were more likely to recall and recognise list items than critical lures. While recall rates were lower than recognition rates, their between-list variability was high for both list items and critical lures (80%–54% and 99%–60% for the recall and recognition of list items, and 47%–0% and 77%–13% for the recall and recognition of critical lures, respectively). The reason why participants recognised list items and critical lures more frequently than recalled them is probably due to the recognition test being applied after the participants had conducted the recall test. As suggested by Roediger & McDermott (1995), this type of experimental design might cause source monitoring errors (Johnson et al., 1993).

Contrary to our expectation, the participants recalled slightly fewer critical lures than unrelated lures (1.32% vs. 1.51%). This might be so because recalled unrelated lures of a certain list may have been presented in—or were even a critical lure of—some other list, making them more readily available for recall as opposed to true unrelated lures. Indeed, further analysis showed that out of the total 169 recalled unrelated lures, 124 were truly unrelated, and 45 were either a list item or a critical lure from one of the previously presented lists. If we consider truly unrelated lures only, then critical lures were recalled more frequently than unrelated lures (1.32% vs. 1.00%), which is in line with our expectations. In contrast, the recognition test consisted of unrelated lures that were completely (or mostly) unrelated to all the lists that were presented to the participant and the participants recognised many more critical lures than unrelated lures (15.6% vs. 2.7%), suggesting that false memory effect was indeed present.

Correlational analysis showed that the participants who recalled or recognised more list items did not necessarily do

so for critical lures (recall and recognition) or unrelated lures (recognition). This suggests that the participants solved the task as instructed ("earnestly and with commitment") and did not simply strive to score as many words as possible, be they true or not. The correlational analysis also showed that the participants who recognised more critical lures also recognised more unrelated lures, which suggests that some participants are more prone than others to false remembering in general. In line with prior research (Anastasi, De Leon, & Rhodes, 2005; Iacullo & Marucci, 2016; Stadler et al., 1999), the results showed that the list item and critical lure recall were correlated with the list item and critical lure recognition on the level of both participants and word lists. This supports the idea that recall and recognition memory are strongly related (Haist et al., 1992) and indicates that the lists used to induce false memories are similarly effective for both recall and recognition.

The comparison of the 18 most effective lists between our study and five similar normative studies showed that the present study demonstrated the lowest rates for critical lures and the highest rates for list items. Although this is not supported by the only other published online DRM study (Murre et al., 2013), one plausible reason for such an outcome is that our results were affected by the online format of the study. Despite our best efforts to control the experimental environment (e.g., by asking participants to find a time and place that would minimise potential interruptions or by excluding from the analysis those who reported having experienced technical issues), the participants might have experienced additional distractions compared to the usual laboratory format which could have undermined the internal validity of the study (Dandurand et al., 2008) and decreased the false memory effect. Moreover, since the participants conducted the study without the physical or virtual presence of the researcher, they might have acted differently from how they normally would (e.g., Friesen et al., 2020). Indeed, varying the demand characteristics (i.e., the effects that are not deliberately caused by the experimental scenario) in the DRM studies has been shown to decrease the false memory effect (Gallo et al., 2001). Taking all this together, a challenge for future studies is to find a way to increase the environmental control in online studies and to explore the effects of the socio-environmental context (e.g., online vs laboratory setting or presence vs absence of a researcher) on the results of the DRM studies.

Turning to RTs, we obtained the same pattern of results for the recall and the (positive responses to the) recognition data. As expected, the RTs of list items and critical lures did not differ from one another, whereas the RTs of unrelated lures were significantly longer than the RTs of both list items and critical lures. These results suggest not only that associatively related information requires less cognitive processing than unrelated information, as predicted by the relatedness effect (Gallo, 2006), but also that closely related false memories (i.e., critical lures) are processed like true memories (i.e., list items), which is in line with the findings by Thomas and Sommers (2005) and Tun et al. (1998).

Considering RTs of negative responses to the recognition test, the results showed the participants were the fastest in rejecting the unrelated lures, slower in rejecting the list items, and the slowest in rejecting the critical lures. While it is not surprising that the participants rejected the unrelated lures the fastest, it is unexpected that the critical lures were rejected at a slower pace compared to the list items. Since positive responses to list items and critical lures were processed similarly fast, one would expect negative responses to be processed similarly fast (Tun et al., 1998). Another reasonable outcome would be that list items, being actually presented, would be rejected slower than critical lures (Jou et al., 2004). Neither was true for our results, which points to the unique character of the critical lures used in the present study.

In terms of RTs of correct responses to the recognition test, the results showed the participants were similarly fast at accepting the list items and rejecting the unrelated lures, but significantly slower at rejecting the critical items, which agrees with prior research (Atkins & Reuter-Lorenz, 2008). This suggests that only closely related false memories (when they are not mistaken for true memories) require additional processing to be evaluated correctly, which makes the DRM paradigm particularly interesting for studying cognitive inhibition (Howe, 2005).

Correlational tests on RTs provided further evidence for a close resemblance between true and closely related false memories, as well as a dissimilarity between true and unrelated false memories. Namely, for both recall and recognition RTs, list items were associated with critical lures, but not with unrelated lures. In addition, recall RTs of critical lures were not associated with unrelated lures. However, we also observed a mild correlation between recognition RTs of critical lures and unrelated lures, which is inconsistent with the above conclusion and points to the similarity between closely related and unrelated false memories. The comparison between RTs of recall and recognition showed a correlation for list items, but not for critical and unrelated lures, which suggests that the relationship between recall and recognition memory is more complex for RTs than accuracy data. Lastly, we found that critical lure recognition rates were negatively associated with corresponding RTs, which indicates that the participants who responded faster to the onset of critical lures also produced more false memories or *vice versa*.

Together with the fact that the present study demonstrated the highest rate of true memories and the lowest rate of closely related false memories in comparison with five similar normative studies, the negative correlation between time and accuracy of false recognition implies that the participants favoured accuracy at the cost of RTs. Such a speed-accuracy trade-off is not surprising given the fact that we instructed the participants to avoid making mistakes, but we did not instruct them to respond as fast as possible. This finding is also consistent with two-process theories of false memory, such as activation-monitoring theory and FTT, which assume that the error-reducing mechanism is slower than the error-inflating mechanism (Arndt & Gould, 2006). Namely, studies showed that speeded retrieval disrupts the error-reducing mechanism, whereas unspeeded retrieval enhances it (Arndt & Gould, 2006). Put differently, increased retrieval time contributes to more true memories and fewer false memories.

To sum up, we showed that the effect of the DRM paradigm can be extended to the Slovenian language environment and demonstrated that the DRM paradigm can

effectively be employed online. Future studies should explore the relationship between RTs and accuracy measures, the relationship between recall and recognition, as well as the potential effects of the experimental environment on the outcome of the DRM paradigm in more detail. We hope that the provided material, the normative data, and rich exploratory analyses will be of use not only to researchers interested in studying false memories in the Slovenian language but also to the wider field of false memory research.

Acknowledgements

References

Albuquerque, P. B. (2005). Produção de evocações e reconhecimentos falsos em 100 listas de palavras associadas portuguesas [Production of false recalls and recognitions in 100 Portuguese associated word lists]. *Laboratório de Psicologia*, 3, 3–12.

Algina, J., Keselman, H. J., & Penfield, R. D. (2005). An alternative to Cohen's standardized mean difference effect size: A robust parameter and confidence interval in the two independent groups case. *Psychological Methods*, *10*(3), 317–328.

Anastasi, J. S., De Leon, A., & Rhodes, M. G. (2005). Normative data for semantically associated Spanish word lists that create false memories. *Behavior Research Methods*, *37*(4), 631–637.

Anastasi, J. S., Rhodes, M. G., Marquez, S., & Velino, V. (2005). The incidence of false memories in native and non-native speakers. *Memory*, *13*(8), 815–828.

Arndt, J., & Gould, C. (2006). An examination of two-process theories of false recognition. *Memory*, *14*(7), 814–833.

Atkins, A. S., & Reuter-Lorenz, P. A. (2008). False working memories? Semantic distortion in a mere 4 seconds. *Memory and Cognition*, *36*(1), 74–81.

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, *11*(5).

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory*. Oxford University Press.

Brainerd, C. J., Yang, Y., Reyna, V. F., Howe, M. L., & Mills, B. A. (2008). Semantic processing in "associative" false memory. *Psychonomic Bulletin and Review*, *15*(6), 1035–1053.

Cabeza, R., & Lennartson, E. R. (2005). False memory across languages: Implicit associative response vs fuzzy trace views. *Memory*, *13*(1), 1–5.

Coane, J. H., McBride, D. M., Raulerson, B. A., & Jordan, J. S. (2007). False memory in a short-term memory task. *Experimental Psychology*, *54*(1), 62–70.

Dandurand, F., Shultz, T. R., & Onishi, K. H. (2008). Comparing online and lab methods in a problem-solving experiment. *Behavior Research Methods*, *40*(2), 428–434.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*(1), 17–22.

Field, A. P., & Wilcox, R. R. (2017). Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers. *Behaviour Research and Therapy*, *98*, 19–38.

Friedrich, S., Konietschke, F., & Pauly, M. (2019). Resampling-based analysis of multivariate data and repeated measures designs with the R package MANOVA.RM. *R Journal*, *11*(2), 380–400.

Friesen, K. B., Zhang, Z., Monaghan, P. G., Oliver, G. D., & Roper, J. A. (2020). All eyes on you: How researcher presence changes the way you walk. *Scientific Reports*, *10*(1), 1–8.

Gallo, D. A. (2006). *Associative illusions of memory: False memory research in DRM and related tasks*. Psychology Press.

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory and Cognition*, *38*(7), 833–848.

Gallo, D. A., & Roediger, H. L. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language*, *47*(3), 469–497.

Gallo, D. A., Roediger, H. L., & McDermott, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin and Review*, *8*(3), 579–586.

Göz, I. (2005). Word frequency effect in false memories. *Psychological Reports*, *96*(3 Pt 2), 1095–1112.

Haist, F., Shimamura, A. P., & Squire, L. R. (1992). On the relationship between recall and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(4), 691–702.

Horoiță, A., & Opre, A. (2020). False memories: Romanian Deese–Roediger–McDermott lists of words. *Cognition, Brain, Behavior. An Interdisciplinary Journal*, *24*(2), 163–186.

Howe, M. L. (2005). Children (but not adults) can inhibit false memories. *Psychological Science*, *16*(12), 927–931.

Iacullo, V. M., & Marucci, F. S. (2016). Normative data for Italian Deese/Roediger–McDermott lists. *Behavior Research Methods*, *48*(1), 381–389.

Johansson, M., & Stenberg, G. (2002). Inducing and reducing false memories: A Swedish version of the Deese–Roediger–McDermott paradigm. *Scandinavian Journal of Psychology*, *43*(5), 369–383.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28.

Johnson, M. K., & Raye, C. L. (1998). False memories and confabulation. *Trends in Cognitive Sciences*, *2*(4), 137–145.

Jou, J. (2008). Recall latencies, confidence, and output positions of true and false memories: Implications for recall and metamemory theories. *Journal of Memory and Language*, *58*(4), 1049–1064.

Jou, J., Matus, Y. E., Aldridge, J. W., Rogers, D. M., & Zimmerman, R. L. (2004). How similar is false recognition to veridical recognition objectively and subjectively? *Memory and Cognition*, *32*(5), 824–840.

Kent, G. H., & Rosanoff, A. J. (1910). A study of association in insanity. *American Journal of Psychiatry*, *67*(1), 37–96.

Lee, Y. S., Chiang, W. C., & Hung, H. C. (2008). Lexical association and false memory for words in two cultures. *Journal of Psycholinguistic Research*, *37*(1), 49–58.

Loftus, E. F. (1993). The reality of repressed memories. *American Psychologist, 48*(5), 518–537.

Loftus, E., & Ketcham, K. (1991). *Witness for the defense: The accused, the eyewitness, and the expert who puts memory on trial.* St. Martin's Press.

Mair, P., & Wilcox, R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods, 52*(2), 464–488.

Marmolejo, G., Diliberto-Macaluso, K. A., & Altarriba, J. (2009). False memory in bilinguals: Does switching languages increase false memories? *American Journal of Psychology, 122*(1), 1–16.

McDermott, K. B., & Watson, J. M. (2001). The rise and fall of false recall: The impact of presentation duration. *Journal of Memory and Language, 45*(1), 160–176.

Miyaji, Y., & Yama, H. (2002). Making Japanese lists which induce false memory at high probability for the DRM paradigm. *The Japanese Journal of Psychonomic Science, 21*(1), 21–26.

Murre, J. M. J., Janssen, S. M. J., Rouw, R., & Meeter, M. (2013). The rise and fall of immediate and delayed memory for verbal and visuospatial information from late childhood to late adulthood. *Acta Psychologica, 142*(1), 96–107.

Payne, J. D., Nadel, L., Allen, J. J. B., Thomas, K. G. F., & Jacobs, W. J. (2002). The effects of experimentally induced stress on false recognition. *Memory, 10*(1), 1–6.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203.

R Core Team. (2022). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803–814.

Russell, W. A., & Jenkins, J. J. (1954). *The complete Minnesota norms for responses to 100 words from the Kent–Rosanoff Word Association Test* (Tech. Rep. No. 11, Contract N8 ONR 66216, Office of Naval Research). University of Minnesota, Twin Cities Campus.

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: Remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1481), 773–786.

Stadler, M. A., Roediger, H. L., & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition, 27*(3), 494–500.

Stein, L. M., da Fonte Feix, L., & Rohenkohl, G. (2006). Avanços metodológicos no estudo das falsas memórias: Construção e normatização do procedimento de palavras associadas [Methodological advances in false memory research: Development and norming of DRM procedure]. *Psicologia: Reflexão e Crítica, 19*(2), 166–176.

Thomas, A. K., & Sommers, M. S. (2005). Attention to item-specific processing eliminates age effects in false memories. *Journal of Memory and Language, 52*(1), 71–86.

Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology and Aging, 13*(2), 230–241.

Ulatowska, J., & Olszewska, J. (2013). Creating associative memory distortions—A Polish adaptation of the DRM paradigm. *Polish Psychological Bulletin, 44*(4), 449–456.

Zeelenberg, R., & Pecher, D. (2002). False memories and lexical decision: Even twelve primes do not cause long-term semantic priming. *Acta Psychologica, 109*(3), 269–284.

Appendix A. The Slovenian DRM Lists

The 36 Slovenian DRM lists that were used in the present study in alphabetical order of critical lures. English translations are given in the parentheses and mean rates for recall and recognition of critical lures as well the number of the corresponding set of lists in the square brackets. The list items are sorted in descending order of association power to the critical lure.

BOLEZEN (SICKNESS) [.00/.40; set 1]	CVET (BLOSSOM) [.00/.37; set 3]	DEKLE (GIRL) [.10/.60; set 2]	DOLGO (LONG) [.00/.20; set 3]
zdravje (health)	roža (flower)	fant (guy)	kača (snake)
rak (cancer)	čebela (bee)	punca (girl)	pot (path)
smrt (death)	dišeč (fragrant)	mladost (youth)	vrv (rope)
gripa (flu)	pomlad (spring)	lepo (beautiful)	čas (time)
vročina (fever)	marjetica (daisy)	kitke (braids)	palica (stick)
virus (virus)	prašnik (anther)	obleka (dress)	meter (meter)
slabost (nausea)	travnik (meadow)	nežnost (tenderness)	čakanje (waiting)
prehlad (cold)	vonj (smell)	deček (boy)	potovanje (travel)
kašelj (cough)	šopek (bouquet)	kmetica (farmer)	predavanje (lecture)
termometer (thermometer)	roza (pink)	prijateljica (friend)	ponedeljek (Monday)
slabo počutje (malaise)	orhideja (orchid)	naivnost (naivety)	leto (year)
starost (oldness)	tulipan (tulip)	moje (mine)	študij (studies)
angina (angina)	alergija (allergy)	najstnica (teenager)	vezenje (embroidery)
ošpice (measles)	rastlina (plant)	Mojca (Mojca)	trak (tape)
depresija (depression)	mak (poppy)	nasmeh (smile)	trajanje (duration)

GLADKO (SMOOTH) [.37/.67; set 2]	GLASBA (MUSIC) [.07/.23; set 1]	GLAVA (HEAD) [.17/.43; set 1]	GLOBOKO (DEEP) [.00/.43; set 2]
led (ice)	note (notes)	možgani (brain)	morje (sea)
hrapavo (coarse)	kitara (guitar)	lasje (hair)	jama (cave)
koža (skin)	klavir (piano)	telo (body)	plitvo (shallow)
površina (surface)	violina (violin)	pamet (sense)	brezno (chasm)
miza (table)	sprostitev (relaxation)	obraz (face)	jezero (lake)
svila (silk)	ples (dance)	vrat (neck)	ocean (ocean)
spolzko (slippery)	pesem (song)	misli (thoughts)	grlo (throat)
grobo (rough)	zvok (sound)	oči (eyes)	dno (bottom)
maslo (butter)	Mozart (Mozart)	lobanja (skull)	prepad (gulf)
papir (paper)	slušalke (headphones)	boli (hurts)	dihanje (breathing)
tla (floor)	rock (rock)	nos (nose)	padec (fall)
moka (flour)	radio (radio)	človek (human)	pogovor (conversation)
testo (dough)	zabava (party)	trup (torso)	potapljanje (diving)
ravno (straight)	melodija (melody)	težka (heavy)	bazen (pool)
marmor (marble)	instrument (instrument)	znanje (knowledge)	glas (voice)

GORA (MOUNTAIN) [.07/.40; set 3]	IGLA (NEEDLE) [.17/.63; set 1]	JEZA (ANGER) [.30/.67; set 2]	KRATKO (SHORT) [.13/.23; set 3]
Triglav (Triglav)	šivanje (sewing)	bes (fury)	ravnilo (ruler)
sneg (snow)	bolečina (pain)	čustvo (emotion)	krilce (skirt)
vrh (top)	nit (thread)	kričanje (screaming)	jedrnato (pithily)
višina (height)	droge (drugs)	žalost (sadness)	življenje (life)
hrib (hill)	injekcija (injection)	agresija (aggression)	besedilo (text)
skale (rocks)	ostro (sharp)	prepir (quarrel)	odgovor (answer)
napor (effort)	sukanec (thread)	srd (wrath)	hitro (quick)

kamen (stone)	vbod (prick)	slaba volja (bad mood)	dan (day)
pohod (hike)	pik (sting)	nasilje (violence)	malo (little)
dolina (valley)	cepivo (vaccine)	udarec (punch)	stavek (sentence)
Everest (Everest)	seno (hay)	sovraštvo (hatred)	hlače (pants)
Alpe (Alps)	narkoman (drug addict)	nemir (restlessness)	črta (line)
Šmarna (Šmarna)	heroin (heroin)	razočaranje (disappointment)	vrvica (string)
plezanje (climbing)	blago (fabric)	trma (stubbornness)	vikend (weekend)
gozd (forest)	krvodajalec (blood donor)	nezadovoljstvo (dissatisfaction)	svinčnik (pencil)

LEV (LION) [.10/.13; set 1]	LUNA (MOON) [.13/.30; set 1]	MEHKO (SOFT) [.10/.70; set 3]	METULJ (BUTTERFLY) [.17/.23; set 1]
griva (mane)	noč (night)	trdo (hard)	krila (wings)
kralj (king)	zvezde (stars)	blazina (cushion)	barve (colors)
žival (animal)	mesec (month)	odeja (blanket)	lepota (beauty)
horoskop (horoscope)	sonce (sun)	vzglavnik (pillow)	cekinček (copper butterfly)
rjovenje (roaring)	polna (full)	puhasto (fluffy)	gosenica (caterpillar)
moč (power)	sije (shines)	mačka (cat)	letenje (flying)
tiger (tiger)	vesolje (space)	toplo (warm)	svoboda (freedom)
mogočnost (mightiness)	krajec (crescent)	deka (blanket)	pisan (colourful)
Afrika (Africa)	skrivnost (secret)	vata (cotton wool)	buba (chrysalis)
savana (savannah)	okrogla (round)	prijetno (pleasant)	rumen (yellow)
safari (safari)	satelit (satellite)	pliš (plush)	let (flight)
ponos (pride)	nespečnost (insomnia)	pero (feather)	poletje (summer)
kletka (cage)	mrk (eclipse)	perje (plumage)	krhkost (fragility)
divjina (wilderness)	mlaj (new moon)	pena (foam)	vešča (moth)
zver (beast)	sanjarjenje (daydreaming)	puding (pudding)	preobrazba (transformation)

MOŠKI (MAN) [.03/.37; set 3]	OKNO (WINDOW) [.33/.73; set 2]	OTROK (CHILD) [.10/.47; set 1]	OVCA (SHEEP) [.07/.60; set 2]
ženska (woman)	pogled (view)	starši (parents)	volna (wool)
spol (gender)	steklo (glass)	veselje (joy)	bela (white)
brada (beard)	vrata (door)	dojenček (baby)	koza (goat)
oče (father)	svetloba (light)	majhen (small)	čreda (herd)
partner (partner)	svet (world)	igra (game)	volk (wolf)
penis (penis)	zavese (curtains)	mama (mum)	pašnik (pasture)
seks (sex)	narava (nature)	ljubezen (love)	pastir (shepherd)
oseba (person)	odprto (open)	družina (family)	jagnje (lamb)
mišice (muscles)	zrak (air)	vrtec (kindergarten)	kmetija (farm)
privlačnost (attraction)	veliko (large)	jok (crying)	dlaka (fur)
brki (mustache)	zunaj (outside)	odrasli (adults)	sledenje (tracking)
gospod (sir)	šipa (windowpane)	nedolžnost (innocence)	bacek (ram lamb)
dedec (fellow)	pokrajina (landscape)	šola (school)	kožuh (fur)
ego (ego)	prozorno (transparent)	odgovornost (responsibility)	pulover (pullover)
brat (brother)	okvir (frame)	smeh (laughter)	kodri (curls)

PAJEK (SPIDER) [.10/.27; set 1]	RDEČE (RED) [.17/.57; set 3]	REKA (RIVER) [.13/.50; set 3]	SADJE (FRUIT) [.03/.23; set 2]
mreža (web)	kri (blood)	voda (water)	jabolko (apple)
fobija (phobia)	črno (black)	Drava (Drava)	banana (banana)

noge (legs)
gnus (disgust)
tarantela (tarantula)
žuželka (bug)
kosmat (hairy)
suha južina (daddy-long-legs)
kot (corner)
osem (eight)
muha (fly)
grd (ugly)
ogabno (disgusting)
dlačice (hairs)
insekt (insect)

vino (wine)
srce (heart)
vrtnica (rose)
ustnice (lips)
menstruacija (menstruation)
šminka (lipstick)
strast (passion)
bik (bull)
komunizem (communism)
živo (live)
češnje (cherries)
kapica (riding hood)
pesa (beet)

Sava (Sava)
potok (stream)
modra (blue)
tok (flow)
Soča (Soča)
most (bridge)
Mura (Mura)
Krka (Krka)
struga (riverbed)
šumenje (murmur)
vijuga (meander)
Sora (Sora)
riba (fish)

zelenjava (vegetables)
hruška (pear)
jagoda (strawberry)
pomaranča (orange)
grozdje (grapes)
mango (mango)
ananas (pineapple)
vitamini (vitamins)
lubenica (watermelon)
košara (basket)
skleda (bowl)
svežina (freshness)
kivi (kiwi)

SIR (CHEESE) [.20/.33; set 2]	SLADKO (SWEET) [.03/.37; set 1]	SOL (SALT) [.03/.20; set 3]	SPANJE (SLEEP) [.27/.70; set 3]
luknje (holes)	čokolada (chocolate)	poper (pepper)	postelja (bed)
mleko (milk)	grenko (bitter)	začimba (spice)	sanje (dreams)
edamec (edam)	bombon (candy)	juha (soup)	počitek (rest)
miš (mouse)	sladoled (ice cream)	okus (taste)	mir (peace)
salama (salami)	torta (cake)	natrijev klorid (sodium chloride)	užitek (pleasure)
kruh (bread)	slaščica (dessert)	Piran (Piran)	udobje (comfort)
gavda (gouda)	med (honey)	kristal (crystal)	bedenje (vigil)
pica (pizza)	čaj (tea)	morska (sea)	utrujenost (fatigue)
Jošt (Jošt)	pecivo (pastry)	čips (crips)	dremanje (napping)
parmezan (Parmesan)	lizika (lollipop)	meso (meat)	dobro (good)
krava (cow)	sok (juice)	jajca (eggs)	umiritev (assuagement)
sendvič (sandwich)	smetana (cream)	zrna (grains)	smrčanje (snoring)
klobasa (sausage)	piškot (biscuit)	rana (wound)	ležanje (lying)
ementaler (emmenthaler)	krof (doughnut)	jod (iodine)	odmor (break)
toast (toast)	palačinke (pancakes)	testenine (pasta)	jutro (morning)

ŠTEDILNIK (STOVE) [.00/.20; set 1]	STRAH (AFRAID) [.30/.63; set 2]	TRG (SQUARE) [.33/.77; set 3]	ULICA (STREET) [.13/.50; set 1]
kuhanje (cooking)	tema (darkness)	vodnjak (well)	mesto (city)
hrana (food)	duhovi (ghosts)	Prešernov (Prešernov)	cesta (road)
ogenj (fire)	trepet (trembling)	kip (statue)	hiša (house)
kosilo (lunch)	tesnoba (anxiety)	selo (village)	ozka (narrow)
vroče (hot)	pogum (courage)	ljudje (people)	številka (number)
plin (gas)	groza (horror)	spomenik (monument)	svetilka (lamp)
lonec (pot)	krik (scream)	Glavni (Glavni)	dom (home)
toplota (heat)	stiska (distress)	center (downtown)	naslov (address)
pečica (oven)	samota (loneliness)	tlakovci (paving)	pločnik (pavement)
plošča (plate)	bojazen (apprehension)	središče (center)	sprehod (walk)
ponev (pan)	votel (hollow)	vas (country)	luč (light)
kava (coffee)	neprijetnost (discomfort)	prostor (space)	sosedje (neighbours)
posoda (dish)	izpiti (exams)	republike (republic)	slepa (blind)
opeklina (burn)	panika (panic)	stojnice (stalls)	blok (block of flats)
plamen (flame)	stres (stress)	Ljubljana (Ljubljana)	hišna številka (house number)

VISOKO (HIGH) [.03/.37; set 2]	ZDRAVNIK (DOCTOR) [.47/.70; set 2]	ZELJE (CABBAGE) [.13/.53; set 3]	ŽELJA (WISH) [.27/.43; set 2]
nizko (low)	halja (gown)	kislo (sour)	darilo (gift)
stolp (tower)	bolnica (hospital)	solata (salad)	ukaz (order)
drevo (tree)	stetoskop (stethoscope)	repa (turnip)	hrepenenje (longing)
nebo (sky)	doktor (doctor)	zajček (bunny)	rojstni dan (birthday)
nebotičnik (skyscraper)	pomoč (help)	vrt (garden)	cilj (goal)
oblaki (clouds)	pregled (check-up)	sarma (cabbage roll)	uspeh (success)
letalo (plane)	bolnik (sick person)	fuj (yuck)	upanje (hope)
ptica (bird)	poklic (profession)	segedin (segedin goulash)	utrinek (shooting star)
strop (ceiling)	medicina (medicine)	njiva (field)	zvezda (star)
stavba (building)	ordinacija (doctor's office)	jota (Istrian stew)	izpolnitev (fulfillment)
moški (man)	sestra (nurse)	smrad (stink)	božič (Christmas)
streha (roof)	pacient (patient)	pečenice (bratwursts)	sreča (luck)
nedosegljivo (unreachable)	ambulanta (clinic)	klobase (sausages)	potreba (need)
košarka (basketball)	specialist (specialist)	jed (food)	trud (effort)
lestev (ladder)	osebni (personal)	kitajsko (Chinese)	motivacija (motivation)

Appendix B. Unrelated Words Used in the Recognition Test

The 72 unrelated words that were used in the recognition test of the present study. English translations are given in the parentheses. The items are sorted in alphabetical order. Some of the unrelated words that we used in the recognition test are similar (or even identical) to the critical lures and list items of the word lists that were presented to the participants (see Appendix A), but, importantly, no such similarity/identity occurred within a given set of lists and its corresponding group of participants. Hence this could not have affected the study's outcome.

Set 1	Set 2	Set 3
avtoriteta (authority)	Amerika (America)	gnezdo (nest)
delo (work)	copat (slipper)	grb (coat-of-arms)
dim (smoke)	čevelj (shoe)	jastreb (hawk)
dolgočasno (boring)	duša (soul)	kremplji (claws)
hoja (walk)	gore (mountains)	letalec (airman)
kazen (punishment)	kladivo (hammer)	Nemčija (Germany)
migrant (migrant)	kljun (beak)	orel (eagle)
nežno (tender)	material (material)	orlovo gnezdo (eagle's nest)
omama (intoxication)	noga (leg)	perspektiva (perspective)
orel (eagle)	oblak (cloud)	planine (mountains)
orodje (tool)	pomembno (important)	plenilec (predator)
pest (fist)	ponoči (at night)	prostost (liberty)
počasi (slow)	prijem (grip)	ptica (bird)
praznovanje (celebration)	ravnanje (treatment)	Rim (Rome)
pretep (fight)	rezanje (cutting)	sokol (falcon)
soba (room)	stopalo (foot)	sova (owl)
sodišče (court)	škarje (scissors)	strela (lightning)
strela (lightning)	tat (thief)	svoboda (freedom)
šef (boss)	težave (problems)	ujeda (bird of prey)
temno (dark)	težko (difficult)	veličastnost (magnificence)
tkanina (fabric)	tobak (tobacco)	velik (large)
uho (ear)	tiho (quiet)	velika ptica (big bird)
ustvarjanje (creating)	zeleno (green)	ZDA (USA)
žebelj (nail)	žeja (thirst)	žival (animal)

Appendix C. Dealing with Errors in the Material and Problematic Responses on the Recognition and Recall Tests

We excluded one list (critical lure *dekle* [*girl*]) from the analysis since it accidentally appeared to have 16 instead of 15 words on it. For the sake of completeness, we still present the recall and recognition rates for the list (ignoring the possible occurrence of the 16th word).

One word (*moški* [*man*]) occurred both as a list item in one list and as a critical lure in another list. This, however, did not introduce any bias to the results, as the list with the critical lure *moški* was part of the third set and the list that contained the word *moški* was part of the second set of lists, and the two sets were given to two different groups of participants.

Two words (*vbod* [*prick*] and *bik* [*bull*]) from two different lists were uttered in such a way by the speaker that one could easily mix them with another word (*pot* [*path*] and *pik* [*sting*], respectively). Indeed, several participants recalled the words in their correct version (7 for *vbod* and 2 for *bik*), however, many participants recalled the words in their alternative version (8 for *pot* and 13 for *pik*) instead. We treated these as correctly recalled words (list items). As the words *vbod* and *bik* (but not *pot* and *pik*) were additionally presented in the recognition test, most of the participants who had heard and recalled the word *pot* and *bik*, we presume, naturally responded with the *N* key, indicating that the word was not presented on the list. Since we cannot know whether they would have recognized the word had they heard it correctly, we decided to exclude the recognition data for the words *vbod* and *bik* altogether. This change did not affect the overall pattern in the statistical analyses.

The participants made several typos when recalling words. For instance, some participants could not use the letters *š*, *č*, or *ž* and could not write certain words correctly (e.g., *moški*), and some had issues with the *Y* and the *Z* key being switched, leading to obvious typos, such as *ljubeyen*, when the word *ljubezen* [*love*] was presented on the list. To prevent wrong attributions to the recalled words when using automated analysis (e.g., marking the recalled item as an unrelated lure, when it should have been marked as a list item), we corrected such occurrences manually.

Finally, some participants recalled words that had the same meaning as the presented ones, but differed subtly in terms of form (e.g., *razočaranje* and *razočaranost* both mean the state of *disappointment*, however, only *razočaranje* was included in the list). We marked such recalled words as correct.

Appendix D. Descriptive Statistics of Accuracy Data for Each Set of Lists

		Accuracy data					
		Recall			Recognition		
Set of lists		1	2	3	1	2	3
List items	N	30	30	30	30	30	30
	M	.68	.70	.69	.77	.77	.82
	SD	.14	.10	.09	.12	.11	.09
	Range	.57	.51	.43	.40	.45	.34
	Skew.	−0.22	0.65	0.12	−0.25	−0.64	−0.04
	Kurt.	−0.67	0.95	−0.17	−1.18	−0.14	−0.64
Critical lures	N	30	30	30	30	30	30
	M	.10	.22	.12	.35	.53	.46
	SD	.10	.17	.09	.25	.27	.18
	Range	.33	.55	.33	.92	1.00	.67
	Skew.	0.92	0.30	0.72	0.49	−0.16	−0.20
	Kurt.	0.00	−1.04	−0.26	−0.92	−0.74	−0.87
Unrelated lures	N				30	30	30
	M				.07	.10	.06
	SD				.09	.06	.09
	Range				.33	.27	.33
	Skew.				1.18	0.59	1.34
	Kurt.				0.42	0.85	1.24

Appendix E. Descriptive Statistics of RTs for Each Set of Lists

		RTs								
		Recall			Recognition (positive)			Recognition (negative)		
Set of lists		1	2	3	1	2	3	1	2	3
List items	*N*	30	30	30	30	30	30	30	30	30
	M	4186.9	3934.7	3825.8	1714.5	1636.7	1345.0	2310.3	2429.9	1978.6
	SD	1256.9	1666.8	1220.5	755.0	948.5	310.6	1287.3	2410.1	749.0
	Range	6303.7	8054.7	4653.2	3073.4	4748.4	1401.2	5178.4	12666.3	2957.6
	Skewness	1.03	2.11	0.96	1.35	2.57	0.57	1.63	3.35	0.73
	Kurtosis	1.84	4.74	0.03	1.18	7.34	0.10	1.94	12.52	−0.09
Critical lures	*N*	20	24	24	28	29	30	30	28	30
	M	9723.8	5368.5	3098.4	1693.2	1452.2	1768.3	2966.3	4394.0	2278.9
	SD	26167.7	4127.9	3213.8	837.4	810.5	1559.0	1691.1	6501.1	677.9
	Range	117993.3	17243.0	11732.8	3039.4	3969.7	8575.5	7544.6	33083.5	3004.5
	Skewness	3.61	1.65	2.16	1.04	2.52	3.92	1.51	3.56	0.23
	Kurtosis	12.16	2.56	3.67	0.19	6.98	16.31	2.43	13.13	−0.27
Unrelated lures	*N*	24	22	23	14	26	14	30	30	30
	M	9646.7	10724.3	7479.4	2750.0	2503.3	2353.0	1526.2	1674.4	1456.8
	SD	6579.3	19647.8	6391.3	1953.5	2544.0	771.5	588.7	873.7	380.3
	Range	22129.2	70368.0	23516.0	7496.2	11347.4	2682.8	2755.6	3589.5	1343.4
	Skewness	0.70	2.49	1.32	1.33	2.48	−0.11	1.86	1.62	0.55
	Kurtosis	−0.78	4.73	0.85	1.47	5.90	−0.86	3.64	1.99	−0.77