Cognition and Neurosciences

“To-be-forgotten” statements become less true: Memory processes involved in selection and forgetting lead to truthfulness changes of ambiguous sentences

ANA SOFIA SANTOS, TÂNIA RAMOS, LEONEL GARCIA-MARQUES and PAULA CARNEIRO

Faculty of Psychology, University of Lisbon, Alameda da Universidade, 1649-013, Lisboa, Portugal


What happens when people try to forget something? What are the consequences of instructing people to intentionally forget a sentence? Recent studies employing the item-method directed forgetting paradigm have shown that to-be-forgotten (TBF) items are, in a subsequent task, emotionally devaluated relative to to-be-remembered (TBR) items, an aftereffect of memory selection (Vivas, Marful, Panagiotidou & Bajo, 2016). As such, distractor devaluation by attentional selection generalizes to memory selection. In this study, we use the item-method directed forgetting paradigm to test the effects of memory selection and inhibition on truth judgments of ambiguous sentences. We expected the relative standing of an item in the task (i.e., whether it was instructed to be remembered or forgotten) to affect the truthfulness value of that item, making TBF items less valid/truthful than TBR items. As predicted, ambiguous sentences associated with a “Forget” cue were subsequently judged as less true than sentences associated with a “Remember” cue, suggesting that instructions to intentionally forget a statement can produce changes in the validity/truthfulness of that statement. To our knowledge, this is the first study to show an influence of memory processes involved in selection and forgetting on the perceived truthfulness of sentences.

Key words: Item-method directed forgetting, memory selection and inhibition, intentional forgetting, judgments of truthfulness, illusions of truth.

Ana Sofia Santos, Faculdade de Psicologia, Universidade de Lisboa. Alameda da Universidade 1649-013 Lisbon, Portugal. Tel: +351 964309241; e-mail: sosantos@fp.ul.pt

INTRODUCTION

The processes of selection and attentional inhibition have been widely studied (Bjork & Murray, 1977; Estes, 1972, Grison, Kessler, Paul, Jordan & Tipper, 2004), but only recently have researchers demonstrated how selective attention is able to influence judgments in an emotional evaluation task (Raymond, Fenske & Tavassoli, 2003). These studies have shown that selection and inhibition responses involved in selective attention tasks are able to influence the affective value of exemplars (for a review see Fenske & Raymond, 2006). For example, Raymond and colleagues (2003) explored the effect of differential prior attention state on targets and distractors by presenting participants with pairs of patterns, one composed of circles and another of squares, and asking them to indicate the location of one of the patterns (i.e., the target) on the screen. Results showed that previously ignored items (i.e., distractors) were evaluated more negatively than targets or new items. In other words, an affective-devaluation effect of distractors was observed (Fenske & Raymond, 2006; Frischen et al., 2012; Raymond et al., 2003; Raymond, Fenske & Westoby, 2005). To explain this phenomenon, Raymond and colleagues (2005) proposed the devaluation-by-inhibition hypothesis: the selection of a target object involves the active inhibition of competing distractors, thus reducing interference with the task-relevant response.

The effects of selection and inhibition responses are not limited to changes in the affective value of items; they also influence other types of judgments, such as typicality judgments (Ramos, Garcia-Marques, Santos & Carneiro, 2015). In fact, recent evidence suggests that selection and inhibition responses affect which exemplars are considered more or less typical of a category – that is, the graded structure of categories (Ramos et al., 2015). The authors tested this prediction with a Go-NoGo task with male and female faces. Results showed that faces associated with a selection (Go) response were judged more typical of their gender category than faces associated with an avoidance (NoGo) response. One possible explanation for these results is that selected faces are tagged as relevant and important and are subsequently regarded as the best exemplars of their category. Interestingly, Raymond and colleagues (2005) made a very similar proposal regarding the attentional distractor devaluation effect. An inhibitory tag would be assigned to the distractor’s representation. Thus, when re-encountered, the inhibitory trace would be reinstated, and the stimulus would be rendered as less emotionally significant.

These selective mechanisms have been suggested to operate in very similar ways in extra attentional fields as well, such as in memory. Researchers have proposed that in paradigms such as the directed forgetting paradigm (Bjork, 1989; Fawcett & Taylor, 2008; Ludowig, Möller, Bien, Münte, Elger & Rosburg, 2010) the better recall/recognition of targeted information implies the active inhibition of distractors. If that idea were correct, it would be conceivable to hypothesize that the distractor devaluation by attentional selection would extend to memory selection. Vivas et al. (2016) recently investigated this connection using the item-method directed forgetting paradigm (Woodward & Bjork, 1971). As predicted, memory processes, such as those involved in intentional forgetting, influenced subsequent emotional evaluations: intentionally forgotten items were subsequently emotionally devaluated (Vivas et al., 2016). As such, distractor
devaluation by attentional selection generalizes to memory selection.

Our aim is to extend these effects of memory selection to judgments of truthfulness. Namely, we investigate if contributions of memory selection and intentional forgetting affect the truth-value of ambiguous statements. This idea derives from the general principle that selection and inhibition are likely to influence the representation of items. We argue that those changes are not limited to the affective value of items, as has been previously investigated, but extend to their relevance (see, Ramos et al., 2015). Specifically, items that are selected as important to our goals are likely to be tagged in memory as relevant and important. In opposition, items that are constantly ignored or purposely forgotten will be tagged as irrelevant or unimportant. These consequences go far beyond the changes in affective value and are probably highly adaptive and crucial for action planning. Items that were previously relevant have a high probability of being relevant in the future, whereas intentionally forgotten items have a high probability of staying irrelevant. In this sense, it is conceivable that, under the assumption that participants mistake relevance for truth, one of the consequences of memory selection would be change in the validity/truth-value of ambiguous statements according to their goal-relevancy. Such an outcome would enable the system to more easily prioritize relevant information and would limit interference caused by competing non-relevant information.

INTENTIONALLY FORGETTING DISTRACTORS IN SOME MEMORY TASKS CLOSELY MIRRORS SUPPRESSING DISTRACTORS IN ATTENTIONAL TASKS

What are the consequences of instructing people to intentionally forget a sentence? To examine these questions, researchers have used the directed forgetting paradigm. In item-method directed forgetting (Woodward & Bjork, 1971), one variant of the directed forgetting paradigm we will focus on, participants are typically presented with single items, each followed by an instruction to remember (TBR condition) or forget (TBF condition). Numerous studies have shown that, on a subsequent memory test, participants typically demonstrate impaired memory for TBF items compared to TBR items (see MacLeod, 1998, for a review), an outcome called the ‘directed forgetting effect’.

Furthermore, when the item-method directed forgetting paradigm is employed, TBF items are subsequently emotionally devaluated relative to TBR items (Vivas et al., 2016). This result has been interpreted as an aftereffect of memory selection (Vivas et al., 2016). In fact, the assumption is that TBF items are actively suppressed, and the forgetting, and consequently inhibition, directed to those items causes emotional devaluation (Vivas et al., 2016).

The underlying framework in these studies is that memory selection raises or brings up a byproduct in the directed forgetting effect. Along with the selection of TBR items comes the forgetting of TBF items, the competing information (Vivas et al., 2016). Some authors (Tipper, 1992) have suggested that such selective and inhibitory mechanisms act in coordination, such that later recognition/recall of targeted information is achieved by the active suppression of distractor information (Fawcett & Taylor, 2008; Ludowig et al., 2010). In this way, memory involves processes able to prioritize goal-directed behavior (Ramos et al., 2015; Vivas et al., 2016).

The occurrence of directed forgetting in the item-method paradigm has long been attributed to processes that enhance the rehearsal/learning of TBR items (Bjork, 1989; Woodward, Bjork & Jongeward, 1973) and not to processes that discourage the encoding or retrieval of TBF items. However, some researchers have suggested that inhibition of these items is indeed involved (e.g., Geiselman & Bagheri, 1985; MacLeod, 1989; Weiner & Reed, 1969, see also Anderson & Hanslmayr, 2014, for a recent review). Recently, Lee, Lee, and Fawcett (2013) used a directed forgetting paradigm to explore the effect of intentional forgetting on color-naming interference within a Stroop-related task. In their studies, Chinese words were presented individually, each followed by an instruction to remember or forget. Across all experiments, participants recalled more to-be-remembered (TBR) than to-be-forgotten (TBF) study words. Moreover, color-naming interference was reduced for repeated TBF words relative to repeated TBR words. That is, TBF items were found to produce less interference on the secondary color-naming task than either repeated TBR items or novel items presented following a “forget” instruction. These findings provide further evidence for the suppression of TBF items at encoding, using a baseline against which to evaluate interference during the word trials. Moreover, an explanation based purely on selective rehearsal would not predict the reduction in color-naming interference for TBF items.

In the current study, we also decided to employ the item-method directed forgetting paradigm because inhibition may be implemented at the level of the individual items. Based on the emotional devaluation effect due to memory selection (Vivas et al., 2016), we expect the relative standing of an item in a task (i.e., whether it was instructed to-be-remembered or forgotten) to affect the truth-value of that item, making TBF items appear less valid/truthful than TBR items.

THE PRESENT STUDY

We aim to test whether judgments of truthfulness are influenced by instructions inducing the selection or forgetting of items through a directed-forgetting paradigm. In item-method directed forgetting, half of the items are followed by a remember instruction, and half are followed by a forget instruction. The order of presentation of items followed by each instruction is random, and participants are told prior to the study their memory will be tested only on those items that were followed by a remember instruction. Typically, items followed by a remember instruction are better recalled than items followed by a forget instruction (for a review, see MacLeod, 1998).

We tested the truth-ambiguous sentences used by Garcia-Marques et al. (2015), in a directed-forgetting paradigm. In our study, participants saw a list of truth-ambiguous sentences, presented in a random order. Some sentences were followed by the word “Remember,” and others were followed by the word “Forget.” With respect to the main purpose of our study, at test, we asked for a judgment of truth of the old sentences and new sentences.

Previous research on truth judgments has shown that repeated statements are afforded higher ratings of truthfulness or are judged
to be true with a higher probability than new statements (e.g., Hasher, Goldstein and Toppino, 1977). By now, this effect has been replicated over 50 times (Dechêne, Stahl, Hansen & Wänke, 2010), demonstrating that the vulnerability of truth judgments to manipulations, such as repeated exposure, is well established.

We propose that the type of an item in a task (i.e., whether it is instructed to-be-remembered or forgotten) should also influence the judgments of truthfulness of those items, when the potential illusion of truth elicited by repetition is kept the same across items. Specifically, we predict that, beyond the typical repetition-induced truth effect, to-be-forgotten ambiguous sentences will be subsequently judged as less valid/truthful than to-be-remembered ambiguous sentences. Such results would mean that intentional forgetting, and consequently inhibition, affects judgments of truth. Furthermore, these effects would go beyond the typical “illusion of truth” effect due to repetition, as they would not depend on the frequency of presentation (i.e., participants would have been exposed to “Remember” and “Forget” sentences with equal frequency). We have no a priori predictions concerning the real status of the statements (whether the statement is actually true or false). Even though true sentences could eventually be judged as significantly more probably true than false sentences, we do not expect this factor to qualify the predicted results for the type-of-item factor.

METHOD

Participants

Seventy-five students from the University of Lisbon (54 women, M_age = 23.28 years, range = 18–51 years) participated in exchange for course credit.

Design

A 2 (status of the statement: true vs. false) × 3 (type of item: remember vs. forget vs. new) within-subjects experimental design was used. The main dependent variable was the average rated truth of the different sentences. We also measured the mean recognition rate as a function of type of item.

Materials

We selected, and used without adaptation, 92 ambiguous sentences with respect to their truth-value from a previous pretest conducted by Garcia-Marques and Silva (unpublished manuscript). See Garcia-Marques et al. (2015) and Garcia-Marques, Silva, and Mello (2016) for details. All stimuli were in Portuguese. We selected ambiguous sentences that were equally likely to be considered true or false, according to the results of the pretest (p > 0.05). An example of a statement is, “crocodiles sleep with their eyes open.” We also used the truth status of the statement as an additional selection criterion, that is, if the sentence was actually true or false. Half of the selected statements (46 sentences) were actually true, and half were actually false (46 sentences). The goal was to control for this factor, but we did not expect any interaction between this factor and our main independent variable (i.e., type of item: remember, forget or new).

Procedure

The experiment was run using E-prime software 2.0 (Schneider, Eschman & Zuccolotto, 2002). Instructions stated that a list of sentences would be presented for a later memory test. Some of the sentences would be followed by the word “Remember,” indicating that participants should try to remember those sentences, while other sentences would be followed by the word “Forget,” indicating that participants should try to forget those sentences. Instructions explained that because participants would be tested only on some of the sentences later, it would be a good idea to follow the information provided by the cues.

Forty sentences were then individually presented on the screen, in a random order, each for 5 seconds. Half of the sentences were followed by a “Remember” cue and half were followed by a “Forget” cue. Each cue was presented for 2 seconds, followed by a 500 ms blank screen. Once all sentences had been presented, participants performed a 4-minute distractor task (a game of Tetris) to minimize any recency effects that could minimize or mask the effects of interest. They then began the test phase.

In a surprise test, participants were asked to rate the perceived degree of truth of 60 randomly presented sentences (20 old sentences and 40 new sentences). Half of all the sentences were factually true, while the remaining were factually false. In addition, half of the old sentences were previously associated with a “Remember” cue, and half were previously associated with a “Forget” cue. Instructions stated that some of the sentences had been previously shown, but that this was irrelevant for the current task. The response scale ranged from “1” (Certainly False) to “6” (Certainly True). Both the sentence and the response scale remained on the screen until the participant provided a response. “Forget,” “Remember,” and “New” sentences were counterbalanced according to a Latin square between participants.

Finally, a recognition test was administered consisting of 24 sentences presented one at a time in a random order: 12 old sentences (6 Remember and 6 Forget) and 12 new sentences. The 12 old sentences included in the recognition test had not been presented in the truth evaluation task. Half of all the sentences were factually true, while the remaining were factually false. The sentences that served as old and new in the recognition test were counterbalanced across participants. Participants were told to decide whether each sentence had appeared in the initial study list, or not, regardless of the cue previously associated with that sentence. They were instructed to press the green key (the “I” key covered by a sticker) for old sentences, and to press the red key (the “E” key covered by a red sticker) for new sentences. There was no time limit to respond. After pressing the key, the sentence was cleared and the next sentence appeared on the screen. At the end of the experiment, participants were debriefed and thanked.

RESULTS

The data of one participant were excluded from analyses because the difference between the correct recognition of TBR and TBF sentences was greater than 2.5 standard deviations below the mean. We interpreted this outcome as a sign that the participant had not followed the instructions; probably failing to inhibit the sentences he was instructed to forget. In the analysis of the results, two-tailed tests were employed for new comparisons (i.e., differences in judgments of truthfulness) and one-tailed tests were employed for well-established findings (i.e., differences in recognition memory between TBR and TBF).

Truth evaluation

Judgments of truthfulness were subjected to a 2 (status of the statement: true vs. false) × 3 (type of item: remember vs. forget vs. new) within-subjects ANOVA. The ANOVA revealed a main effect for type of item, F(2, 146) = 30.61, p < 0.001, MSE = 0.81, η^2_p = 0.29 (means are shown in Table 1, left column). New sentences were judged to be less true than sentences previously presented in the study phase, t(73) = 6.32, p < 0.001, d = 0.73,
two-tailed – a replication of the Illusion of Truth effect (Hasher et al., 1977) using the directed forgetting paradigm. Our central hypothesis was that to-be-forgotten ambiguous sentences would be judged as less truthful than to-be-remembered ambiguous sentences. A specific planned comparison showed that sentences associated with a “Remember” cue were judged as significantly more truthful than sentences associated with a “Forget” cue, $t(73) = 2.96, p = 0.004, d = 0.35$, two-tailed. These results support the idea that memory processes involved in the selection and intentional forgetting of ambiguous sentences influence how these sentences were subsequently judged in terms of truthfulness.

The ANOVA also revealed a main effect of the status of the statements (true vs. false), $F(1, 73) = 19.24, p < 0.001$, $MSE = 0.37$, $\eta^2_p = 0.21$, with true sentences being judged as significantly more probably true ($M = 4.25, SE = 0.13$) than false sentences ($M = 3.99, SE = 0.15$). More importantly, this factor did not qualify the results obtained with the type-of-item factor, as the interaction was non-significant (mean evaluations of truthfulness as a function of type of item and status of the statement are shown in Table 2).

**Table 2. Mean evaluations of truthfulness as a function of type of item and status of the statement**

<table>
<thead>
<tr>
<th>Status of the statement</th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of Item</td>
<td>M (SE)</td>
<td>M (SE)</td>
</tr>
<tr>
<td>Forget</td>
<td>4.32 (0.12)</td>
<td>4.13 (0.13)</td>
</tr>
<tr>
<td>Remember</td>
<td>4.62 (0.11)</td>
<td>4.30 (0.13)</td>
</tr>
<tr>
<td>New</td>
<td>3.79 (0.07)</td>
<td>3.55 (0.08)</td>
</tr>
</tbody>
</table>

**Recognition test**
The mean proportion of “old” responses, that is, the corrected recognition of presented sentences and false recognitions of new sentences, was subjected to a one-way repeated measure ANOVA, with type of item (remember vs. forget vs. new) as the within-subject factor. The main effect of type of item was significant, $F(2, 146) = 755.84, p < 0.001$, $MSE = 0.019$, $\eta^2_p = 0.91$ (means are shown in Table 1, right column). A planned comparison showed a greater proportion of old responses to old sentences than to new sentences, $t(73) = 39.87, p < 0.001, d = 4.66$, one-tailed, confirming that participants were able to discriminate between old and new sentences ($M_{old} = 0.87$ versus $M_{new} = 0.10$).

Our specific theoretical hypothesis was that the to-be-forgotten ambiguous sentences would be less likely to be recognized. A planned one-tailed t-test supported this prediction, $t(73) = 1.85$, $p = 0.034, d = 0.43$. Accuracy was worse for “Forget” ($M = 0.85$) than for “Remember” ($M = 0.89$) sentences. As such, recognition performance indicated a directed-forgetting effect, suggesting that participants followed the instructions to remember or forget the study items.

**GENERAL DISCUSSION**
Recent evidence suggests that the distractor devaluation effect (Fenske et al., 2006; Raymond et al., 2003, 2005) can be observed for intentionally forgotten items (Vivas et al., 2016), extending the effects of selective attention to selective memory processes. Our results converge on these findings, further extending the impact of selective memory processes to judgments of truthfulness. Namely, they indicate that memory processes involved in the selection and intentional forgetting of ambiguous sentences influence how these sentences are subsequently judged in terms of their truthfulness. With respect to truth-value, ambiguous sentences associated with a “Forget” cue were subsequently judged as less true than sentences associated with a “Remember” cue, despite the fact that both types of sentences had been previously presented in the study phase and were judged as more probably true than new sentences — the typical illusion of truth effect (Hasher et al., 1977). Our findings, therefore, are consistent with results typically obtained with item-method directed forgetting and extend the effect of the directed forgetting paradigm to truth judgments.

Moreover, our results support the notion that complex materials such as sentences also lend themselves to directed forgetting. Research investigating item-method-directed forgetting typically uses words as study materials, but directed forgetting effects have also been documented for other stimuli (e.g., pictorial scenes, line drawings, unnamable symbols, visual live-action sequences, single phone numbers, unfamiliar faces, etc.). By using sentences, we have demonstrated item-method-directed forgetting for additional complex materials, which adds to the frequent discussions addressed by this literature.

Significant differences in truth judgments under “Remember” and “Forget” instructions were obtained. Furthermore, a “Forget” cue led to poorer recognition performance than a “Remember” cue or new sentences. However, it remains unclear whether sentences lend themselves easily to a spontaneous stop in elaboration, as occurs with words. It seems unlikely that people rapidly stop mobilizing attentional resources to encode information within the sentence. Most likely, information within sentences followed by either an instruction to remember or forget will be, to a certain degree, encoded and made accessible for future retrieval. Results obtained in this experiment seem to be in line with the expected poorer recognition memory, regarding the preceding “Forget” memory instruction, due to reduced elaborative rehearsal of TBF sentences. However, note that the previous reasoning is not at odds with the rather good recognition memory performance obtained for both type of sentences, regardless of memory instruction. Still, a memory measure more sensitive to “elaborative rehearsal” would have been free recall, which is also frequently used in directed forgetting studies.

Additionally, as we did not include a baseline condition in the directed forgetting task, we are not able to exclude the possibility...
that the difference between TBF and TBR ambiguous sentences was due to an increase in truthfulness judgments for TBR sentences, rather than to a decrease in the judgments for TBF sentences. Future studies could clarify this question by including a baseline condition in which some of the sentences are presented at study without any memory cues.

It is important to note that in our study, recollection of the previous cues is non-diagnostic for their truth-value and thus independent from performance in the truthfulness task. This is in contrast to the procedure used by Skurnik, Yoon, Park and Schwarz (2005), in which memory for the cues is totally diagnostic of the truth evaluation task. In the studies conducted by Skurnik and colleagues (2005), participants were warned that a sentence was true or false during encoding. In fact, the aim of the authors was to investigate the role of memory for this context information (i.e., the warnings that a claim was false or true) in creating the illusion of truth. Under these conditions, when people tried to determine the truth of a claim that felt familiar, memory for the original warning associated with the claim was as important as memory for the claim itself. Thus, remembering that a claim was identified as false helped preventing the illusion of truth of repeated information. Our study differs from Skurnik et al. (2005) because the effect of repeated exposure and recollection are not pitted against one another. Thus, we were able to assess the impact of our manipulation on truth judgments independently of its effect on recollection.

Nevertheless, it is still possible that participants used information about the cues in their judgments, even if the cues are not absolutely diagnostic of the truth value of the sentences. Specifically, the instruction to “remember” may itself be interpreted as meaning that the sentence is relevant, and as a consequence, more likely to be true. In contrast, the instruction to “forget” may be interpreted as a sign of the irrelevancy or falsity of the sentence. If so, there would be no need to postulate inhibitory mechanisms to explain the current findings. This alternative interpretation is consistent with the notion that humans are attentive to conversational cues that indicate the intentions of the communicators. That is, listeners try to decipher what the communicators intend to transmit, beyond what is explicitly stated (Sperber & Wilson, 1987; Schwarz, 1994). Thus, participants could use the following reasoning: if the experimenter tells me to remember, then the information must be relevant; on the other hand, if the experimenter tells me to forget, then the information must be irrelevant. Interestingly, Dittrich and Klauer (2012) argued that affective devaluation effects could be explained by similar mechanisms, and not necessarily by attentional inhibition processes.

However, there are some reasons to favor an explanation based on memory inhibitory mechanisms, in our case. First, instructions in the truth judgment task explicitly stated that, although some of the sentences had been previously shown, that was irrelevant to the judgment task. This instruction should discourage participants of using the connotation of the memory cues to evaluate the truth of the sentences. Second, and more importantly, if participants systematically used the cues (remember and forget) to infer the truth value of the sentences, then one would expect TBF sentences to be rated as more false comparing with new sentences, which was not the case. Although these facts support an inhibition account, it is important for future studies to clarify the underlying mechanisms. One possibility would be to include an initial instruction stating that the pairing between the sentences and the cues is determined randomly by the computer. In that case, it would not make sense to use the cues to infer about the truth value of the sentences.

Finally, there is an aspect of these results relevant to the “truth effect” literature that is also worth considering. Most prior research that has explored the relation between statement repetition and perceived validity has considered repetition to be a driving force in the increased truth judgments. This effect seems to be independent of levels-of-processing manipulations (Begg, Armour & Kerr, 1985), and conscious recollection (Begg, Anas & Farinacci, 1992). Given the robustness of the effect under various contextual conditions, the differences found by uniquely manipulating the instruction associated with the ambiguous sentences seems a novel contribution. Note that the effects on truthfulness we obtained did not depend on the frequency of presentation, as participants were exposed to “to-be-remember” and “to-be-forgotten” sentences with equal frequency.

In sum, our results show that instructions to intentionally forget a statement can produce changes in judgments of truthfulness of that statement, beyond the expected repetition effect on truth judgments. This loss of validity as an aftereffect of ignoring and forgetting information seems adaptive as a way of directing behavior to relevant goals. If goal-irrelevant information were inhibited and tagged as not valid, it would be less likely to interfere with our pursuit of goals, a mechanism that probably supports the constant need to disentangle the essential from the superfluous in decision-making processes.

NOTE

1 Sentence type was counterbalanced within each task, but not between tasks. That is, a set of the sentences from the study phase was selected to be rated in terms of truthfulness and another set was selected to be presented in the recognition test.

REFERENCES


Received 5 August 2016, accepted 7 March 2017